



JOSIP JURAJ STROSSMAYER UNIVERSITY OF OSIJEK

Faculty of Electrical Engineering, Computer Science and Information
Technology Osijek

Petra Radočaj, M. Sc.

Optimization of deep learning methods for pediatric pneumonia classification from chest X-rays

Ph. D. Thesis

Osijek, 2025.

Optimization of deep learning methods for pediatric pneumonia classification from chest X-rays
Unaprjeđenje metoda dubokog učenja za klasifikaciju pneumonije u djece iz radioloških snimki

Petra Radočaj, 2025.

SUPERVISOR

Prof. Goran Martinović, Ph.D.

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, J. J.
Strossmayer University of Osijek, Croatia

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	1
INFORMATION ABOUT THE MENTOR.....	2
INFORMACIJE O MENTORU	4
ABSTRACT.....	6
SAŽETAK.....	7
1. INTRODUCTION.....	8
1.1. GLOBAL BURDEN OF PEDIATRIC PNEUMONIA	8
1.2. CHALLENGES IN DIAGNOSIS AND CURRENT DIAGNOSTIC APPROACHES	9
1.3. DEEP LEARNING IN MEDICAL IMAGING FOR PNEUMONIA DIAGNOSIS	10
1.4. RESEARCH GAP AND MOTIVATION.....	14
1.5. RESEARCH HYPOTHESES	15
1.6. ORIGINAL SCIENTIFIC CONTRIBUTIONS	15
1.7. CONTRIBUTIONS PAPERS	16
2. DISCUSSION.....	17
2.1. METHODOLOGICAL FRAMEWORK AND CONTRIBUTIONS.....	17
<i>2.1.1. Dataset curation and preprocessing</i>	<i>17</i>
<i>2.1.2. Data augmentation and robustness.....</i>	<i>18</i>
<i>2.1.3. Model construction and transfer learning.....</i>	<i>18</i>
<i>2.1.4. Research methodology.....</i>	<i>18</i>
<i>2.1.5. Validation strategies.....</i>	<i>19</i>
<i>2.1.6. Methodological contributions.....</i>	<i>20</i>
2.2. VALIDATION STRATEGIES AND EXPERIMENTAL ROBUSTNESS	20
<i>2.2.1. Stratified validation split.....</i>	<i>21</i>
<i>2.2.2. Evaluation metrics</i>	<i>21</i>
<i>2.2.3. Epoch analysis and learning dynamics.....</i>	<i>23</i>
<i>2.2.4. Robustness considerations.....</i>	<i>23</i>
2.3. THE IMPACT OF ACTIVATION FUNCTION SELECTION	24
2.4. ARCHITECTURAL IMPROVEMENTS FOR FEATURE REPRESENTATION.....	27
2.5. MODEL INTERPRETABILITY AND CLINICAL TRUST	33
2.6. COMPARATIVE PERFORMANCE OF DEEP LEARNING ARCHITECTURES	37
2.7. CONVERGENCE DYNAMICS AND TRAINING STABILITY	42

2.8.	LIMITATIONS, GENERALIZABILITY, FUTURE DIRECTIONS, AND PRACTICAL IMPLICATIONS.....	44
3.	CONCLUSIONS.....	47
	REFERENCES.....	51
	APPENDICES.....	64
	CURRICULUM VITAE	116

ACKNOWLEDGEMENTS

Children are always worth fighting for. With all my heart, I hope to contribute to the field of pediatrics through new methods and perspectives, empowered by artificial intelligence. I truly believe in the progress of machine learning and its meaningful integration into medicine, in partnership with healthcare professionals. After all, no one knows everything, but together, humans and technology can form a remarkable and complementary alliance.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Goran Martinović, Ph.D., who has guided me not only during this doctoral journey but also throughout my earlier studies. His consistent support, encouragement, and insightful mentorship have been instrumental in shaping my academic path. I am deeply thankful for their patience, trust, and the opportunity to grow under their guidance. I would also like to extend my sincere thanks to the Doctoral Committee for their invaluable feedback and dedication.

I owe special thanks to my brother, who introduced me to the world of science. He has been a true role model, someone from whom I have learned deeply and with whom I have shared many meaningful and joyful experiences. His path helped illuminate my own.

To Danijel, thank you for opening doors to the fascinating world of artificial intelligence and for nurturing a strong desire to apply machine learning in practice. I am excited about all the projects we will work on together and grateful for the opportunity to learn from such a forward-thinking environment.

Finally, I am endlessly thankful to my parents, for their unwavering moral and financial support, and for encouraging me every step of the way, from the first day of my studies to the last. With them, I have lived through countless beautiful moments that I will forever treasure.

INFORMATION ABOUT THE MENTOR

Goran Martinović was born in 1969 in Orahovica, Republic of Croatia. In 1996, he graduated from the university study program in Electrical Engineering at the Faculty of Electrical Engineering, J.J. Strossmayer University of Osijek. Since 1996, he has been employed at the same faculty as a junior assistant. In 2000, he received his Master's degree in Computer Science at the Faculty of Electrical Engineering and Computing, University of Zagreb. Since 2000, he has worked as an assistant. In 2004, he earned his PhD in Technical Sciences, field of Computer Science, at the Faculty of Electrical Engineering and Computing, University of Zagreb. During 2004, he worked as a senior assistant, from December 2004 as an assistant professor, from 2009 as an associate professor, from 2013 as a full professor, and since 2017 as a full professor with tenure at FERIT Osijek. Since 2005, he has served multiple terms as Vice-Dean. During 2005 and 2006, he established the doctoral study program and the specialist study program at FERIT Osijek. He has supervised more than 200 defended diploma and bachelor theses, two master's theses, and six doctoral dissertations, and has co-supervised one doctoral dissertation. His scientific and professional work is in the fields of artificial intelligence, data analysis, embedded and distributed computer systems, software engineering, and real-time systems.

During 1996 and 1997, and from 1997 to 2002, he was a collaborator on three projects, and since 2002 a researcher on one project of the Ministry of Science, Education and Sports of the Republic of Croatia. Since 2007, he has been a researcher on one project, and project leader of "Scheduling Procedures in Self-Sustained Distributed Computer Systems" of the Ministry of Science, Education and Sports of the Republic of Croatia. During 2007 and 2008, he was a researcher on a TEMPUS project, and in 2012 and 2013 a leader of an IPA IVC project. From 2007 to 2011, he was an external collaborator-researcher on the PROGRESS project, and from 2010 to 2016 on the RALF 3 project (Swedish Foundation for Strategic Research). Since 2017, he has been a lecturer on the Erasmus KA2 project TEAMSOC21, since 2019 co-leader at FERIT Osijek on the SMARTSOC project, and since 2016 a member of DATACROSS – Center of Excellence for Data Science and Cooperative Systems: RU for Data Science, and since 2017 a researcher on the DATACROSS project. Since 2018, he has been a researcher and leader at FERIT Osijek on the project "Research of Beacons for the Purpose of Building a Mobility Network" within the program "Increase of Development of New Products and Services". Since 2018, he has been a member of the FERIT Osijek project team on the project "ICT in Agricultural Sciences". Since 2019, he has been a collaborator on two CROQF projects. He has participated/is participating in three COST actions: OpenMultiMed, MPM4CPS, and since 2019 in Multi3Generation. Since 2020, he has been a leader at FERIT and a work package leader of the project EUROCC – National Competence Centres in the framework of EuroHPC (European High-Performance Computing Joint Undertaking (JU) under GA No. 951732 and Horizon 2020), and since 2023 leader of the second phase of the same project under the title EuroCC 2 from the Digital Europe call (GA No. 101101903). He has participated in the evaluation of a number of research projects of different calls. Since 1996 until now, he has stayed at several foreign institutions. In 2008, he obtained the diploma of external auditor for quality assurance in higher education from AZVO. He was appointed Chairman of the User Council of CRO NGI, and a member of the Council of CRO NGI. Since 1996, he has stayed at several foreign institutions: Hochschule Bremen, Slovak Technical University Bratislava, Technische Universität Wien, Universität des Saarlandes, Saarbrücken, and

at the International Conference and Research Center for Computer Science Schloss Dagstuhl. From 2008 to 2019, he repeatedly stayed as part of research work on projects at Mälardalen University, Västerås, Sweden.

He is the author or co-author of about 140 scientific and professional papers, and has presented more than 30 communications, mostly at international scientific conferences. He is co-author of six university textbooks. He has delivered two invited lectures at domestic and foreign universities. He has been editor of several proceedings and a book of abstracts of an international conference. He is a member of the editorial board of three journals, as well as a program committee member and reviewer of a number of conferences and journals. In 2009, he launched the *International Journal of Electrical and Computer Engineering Systems*, where he is Editor-in-Chief. He is a member of IEEE and ACM, and founder and chair of the IEEE SMC Chapter, IEEE Croatia Section. Since 2007, he has been a member of the IEEE SMC TC on Distributed Intelligent Systems. He is a member of the Croatian Academy of Engineering, Department of Information Systems. In 2006, he received the IEEE Letter of Appreciation IEEE SMC; in 2008, the IEEE Croatia Section Award for outstanding contribution in engineering education; and in 2019, under his leadership, the IEEE SMC Chapter was awarded the IEEE Outstanding SMC Chapter Award.

He is married. He speaks English and German.

INFORMACIJE O MENTORU

Goran Martinović rođen je 1969. godine u Orahovici, Republika Hrvatska. Godine 1996. diplomirao je na sveučilišnom studiju elektrotehnike Elektrotehničkog fakulteta Sveučilišta J.J. Strossmayera u Osijeku. Od 1996. radio je na istom fakultetu kao mlađi asistent. Magistrirao je 2000. računarstvo na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Od 2000. radio je kao asistent. Doktorirao je 2004. tehničke znanosti, polje računarstvo na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Tijekom 2004. radio je kao viši asistent, od prosinca 2004. kao docent, od 2009. kao izvanredni profesor, od 2013. radi kao redoviti profesor, a od 2017. kao redoviti profesor u trajnom izboru na FERIT-u Osijek. Od 2005. godine u više mandata bio je bio je prodekan. Tijekom 2005. i 2006. ustrojio je doktorski studij, te specijalistički studij na FERIT-u Osijek. Mentor je na preko 200 obranjenih diplomskih i završnih radova, dva magistarska rada i šest doktorskih disertacija, a na jednoj doktorskoj disertaciji bio je sumentor. Znanstveno i stručno djeluje u području računalne inteligencije, analize podataka, ugradbenih i raspodijeljenih računalnih sustava, programskog inženjerstva i sustava za rad u stvarnom vremenu.

Tijekom 1996. i 1997., te od 1997. do 2002. suradnik je na tri projekta, a od 2002. istraživač na jednom projektu MZOS RH. Od 2007. istraživač je na jednom projektu, te voditelj projekta „Postupci raspoređivanja u samoodrživim raspodijeljenim računalnim sustavima“ MZOS RH. Tijekom 2007. i 2008. istraživač je na TEMPUS projektu, a 2012. i 2013. voditelj IPA IVC projekta. Od 2007. do 2011. je vanjski suradnik-istraživač na projektu PROGRESS, a od 2010. do 2016. na projektu RALF 3 (Swedish Foundation for Strategic Research). Od 2017. je predavač na Erasmus KA2 projektu TEAMSOC21, od 2019. suvoditelj na FERIT-u Osijek na projektu SMARTSOC, a od 2016. član je DATACROSS – Center of Excellence for Data Science and Cooperative Systems: RU for Data Science, a od 2017. istraživač na projektu DATACROSS. Od 2018. istraživač je i voditelj na FERIT-u Osijek projekta Istraživanje beacons u svrhu izgradnje mreže kretanja iz programa Povećanje razvoja novih proizvoda i usluga. Od 2018. je član projektnog tima FERIT-a Osijek na projektu ICT u poljoprivrednim znanostima. Od 2019. je suradnik na dva projekta HKO. Sudjelovao je/sudjeluje na tri COST akcije: OpenMultiMed, MPM4CPS, te od 2019. godine na Multi3Generation. Od 2020. godine voditelj je na FERIT-u i voditelj radnog paketa projekta EUROCC - National Competence Centres in the framework of EuroHPC (European High-Performance Computing Joint Undertaking (JU) under GA No. 951732 and Horizon 2020), a od 2023. voditelj druge faze provedbe istog projekta pod nazivom EuroCC 2 iz natječaja Digital Europe (GA No 101101903). Sudjelovao je u vrednovanju niza istraživačkih projekata različitih natječaja. Od 1996. godine do sada, boravio je na više stranih institucija. Godine 2008. stekao je diplomu vanjskog auditora za kvalitetu visokog obrazovanja AZVO. Imenovan je predsjedavajućim Vijeća korisnika CRO NGI, te članom Savjeta CRO NGI. U razdoblju od 1996. godine, boravio je na više stranih institucija: Hochschule Bremen, Slovak Technical University Bratislava, Technische Universität Wien, Universität des Saarlandes, Saarbrücken, te u Međunarodnom konferencijskom i istraživačkom centru za računarstvo Schloß Dagstuhl. Od 2008. do 2019. godine u više navrata boravio je u sklopu istraživačkog rada na projektima na Mälardalen University, Västerås u Švedskoj.

Autor je ili suautor oko 140 znanstvenih i stručnih radova, pri čemu je održao više od 30 priopćenja većinom na međunarodnim znanstvenim skupovima. Suautor je na šest sveučilišnih udžbenika. Održao je dva pozvana predavanja na domaćem i stranom sveučilištu. Urednik je više zbornika i

knjige sažetaka međunarodne konferencije. Član je uređivačkog odbora tri časopisa, te član programskog odbora i recenzent niza konferencija i časopisa. Godine 2009. pokrenuo je *Int. J. of Electrical and Computer Engineering Systems* čiji je glavni suurednik. Član je IEEE i ACM, te osnivač i voditelj IEEE Odjela SMC, IEEE Hrvatska sekcija. Od 2007. član je IEEE SMC TC on Distributed Intelligent Systems. Član je Akademije tehničkih znanosti Hrvatske, Odjel za informacijske sustave. Godine 2006. dobitnik je IEEE Letter of Appreciation IEEE SMC, 2008. dobitnik Nagrade Hrvatske sekcije IEEE za izuzetan doprinos u inženjerskoj edukaciji, a 2019. godine odjel IEEE SMC pod njegovim vodstvom nagrađen je s IEEE Outstanding SMC Chapter Award.

Oženjen je. Govori engleski i njemački jezik.

ABSTRACT

Pneumonia remains a leading cause of morbidity and mortality in children under five, particularly in low- and middle-income countries (LMICs), where limited healthcare access and diagnostic challenges exacerbate the disease burden. Despite being preventable and treatable, pediatric pneumonia accounts for approximately 14% of under-five deaths globally, highlighting the urgent need for accurate and scalable diagnostic solutions. Deep learning, particularly convolutional neural networks (CNNs), has emerged as a promising tool for automated pneumonia classification in chest radiographs, offering potential improvements over traditional clinical and radiological assessments. This study explores the optimization of CNN architectures for pediatric pneumonia classification, focusing on three key aspects: (1) the impact of advanced activation functions (Swish, Mish) compared to traditional ReLU; (2) the integration of multi-scale and strided convolutions to enhance feature representation; and (3) the application of Gradient-weighted Class Activation Mapping (Grad-CAM) to improve model interpretability. Through these approaches, the dissertation contributes to the improvement of CNN architectures adapted for medical image analysis using Swish and Mish activation functions, aiming to reduce information loss during training. Experimental results demonstrate that Mish-activated CNNs achieve superior classification accuracy (up to 97.61%) by preserving gradient flow and reducing information loss during training. Additionally, the integration of multi-scale and atrous convolutions increases the adaptability and performance of CNN architectures, enabling more robust recognition of pneumonia across diverse pediatric chest X-ray datasets. Architectural enhancements, including multi-scale convolutions in models like DenseNet201 and InceptionResNetV2, further improve robustness across diverse pediatric chest X-ray datasets. Additionally, Grad-CAM visualizations align model decisions with clinically relevant regions, fostering trust in AI-assisted diagnostics. This contributes to improving diagnostic reliability and enabling more precise determination of important areas in medical images by visualizing activation maps. Despite these advancements, challenges persist, including dataset limitations, model generalizability, and computational demands in low-resource settings. Future research should focus on multi-institutional data collaboration, domain adaptation techniques, and lightweight model deployment to bridge the gap between AI innovation and clinical implementation.

Keywords: convolutional neural networks; deep learning; feature extraction; medical image analysis; Mish activation function; pediatric pneumonia; transfer learning

SAŽETAK

Pneumonija je i dalje vodeći uzrok morbiditeta i mortaliteta u djece mlađe od pet godina, posebno u zemljama s niskim i srednjim prihodima, gdje ograničen pristup zdravstvenoj zaštiti i dijagnostički izazovi doprinose učestalosti bolesti. Iako se može spriječiti i liječiti, pedijatrijska pneumonija čini oko 14% smrtnih slučajeva u djece mlađe od pet godina, što naglašava hitnu potrebu za preciznim i skalabilnim dijagnostičkim rješenjima. Metode dubokog učenja, posebno konvolucijske neuronske mreže pokazale su obećavajuće rezultate u automatskoj klasifikaciji pneumonije na rendgenskim snimkama prsnog koša, nudeći potencijalnu nadmoć nad dosadašnjim kliničkim i radiološkim metodama. Ova studija istražuje optimizaciju arhitektura konvolucijskih neuronskih mreža za klasifikaciju pedijatrijske pneumonije, s naglaskom na tri ključna aspekta: (1) utjecaj naprednih aktivacijskih funkcija (Swish, Mish) u usporedbi s tradicionalnom ReLU funkcijom; (2) integraciju višeskalnih i konvolucija s povećanim pomakom za poboljšanje reprezentacije značajki; te (3) primjenu metode Gradient-weighted Class Activation Mapping (Grad-CAM) za povećanje interpretabilnosti modela. Eksperimentalni rezultati pokazuju da modeli konvolucijskih neuronskih mreža s Mish aktivacijskom funkcijom postižu veću točnost klasifikacije (do 97.61%) zahvaljujući boljem protoku gradijenta i smanjenom gubitku informacija tijekom treniranja. Ovim je ostvaren prvi doprinos disertacije, unaprjeđenje CNN arhitektura prilagođenih medicinskoj analizi slika uporabom naprednih aktivacijskih funkcija, s ciljem smanjenja degradacije informacija u procesu učenja. Arhitektonske nadogradnje, uključujući višerazinske konvolucije u modelima poput DenseNet201 i InceptionResNetV2, dodatno povećavaju robusnost na različitim skupovima pedijatrijskih rendgenskih snimki. Na taj se način ostvaruje drugi doprinos, povećana prilagodljivost i performanse CNN arhitektura temeljenih na višeskalnim i atrous konvolucijama, što omogućuje prepoznavanje raznolikih patoloških obrazaca u dječjim plućima. Nadalje, Grad-CAM vizualizacije usklađuju odluke modela s klinički relevantnim regijama, potičući povjerenje u AI-pomoćnu dijagnostiku. Time se ostvaruje treći doprinos disertacije, poboljšanje pouzdanosti dijagnostike i preciznije određivanje važnih područja u medicinskim slikama kroz vizualizaciju aktivacijskih mapa, čime se osnažuje klinička primjenjivost sustava. Unatoč napretku, ostaju izazovi poput ograničenosti dostupnih podataka, generalizacije modela i računalnih zahtjeva u resursno ograničenim okruženjima. Buduća istraživanja trebala bi se usredotočiti na međuinstitucionalnu suradnju u prikupljanju podataka, tehnike prilagodbe domena i implementaciju pojednostavljenih modela kako bi se omogućila šira klinička primjena.

Ključne riječi: konvolucijske neuronske mreže; duboko učenje; ekstrakcija značajki; analiza medicinskih slika; Mish aktivacijska funkcija; pedijatrijska pneumonija; prijenosno učenje;

1. INTRODUCTION

1.1. Global burden of pediatric pneumonia

Pneumonia remains one of the most pressing global public health challenges, particularly among children under five years of age [1], [2], [3]. It is a form of acute lower respiratory tract infection that causes inflammation of the alveoli, leading to the accumulation of fluid or pus in the lungs, impaired oxygen exchange, and, if untreated, potentially fatal respiratory failure [4]. Despite being both preventable and treatable, pediatric pneumonia continues to cause significant morbidity and mortality worldwide, disproportionately affecting children in LMICs [5], [6]. Globally, pneumonia affects an estimated 150 million children each year, with the highest prevalence in LMICs [7]. With an annual incidence of 1 in 71 children, South Asia, West Africa, and Central Africa are disproportionately affected. This is due to systemic factors like low vaccination rates, exposure to indoor air pollution, malnutrition, and limited healthcare access [8], [9]. Due to underlying causes in these regions, the prevalence of pediatric pneumonia is high and regularly severe [6]. Within the pediatric population, particularly in those under five years of age, pneumonia is the leading cause of respiratory-related hospital admissions and a significant contributor to overall mortality [10]. As the leading cause of respiratory-related hospitalizations and a primary contributor to under-five mortality, pneumonia accounts for 14% of deaths in this age group, claiming approximately 740,000 young lives annually [11]. Some estimates suggest an even higher toll, ranging between 700,000 and 1 million deaths per year [12], , surpassing the combined mortality of HIV/AIDS, malaria, and measles [13]. While high-income countries report significantly lower mortality rates due to advanced healthcare systems, antibiotic availability, and robust vaccination programs (e.g., against *Streptococcus pneumoniae* and Hib) [8], LMICs struggle with delayed diagnoses and inadequate treatment [14], [15]. Barriers such as poor infrastructure, shortages of trained medical staff, and unreliable diagnostic tools exacerbate outcomes in these regions [6]. It is paramount that this situation should be attended to in its early stages before it develops to more serious conditions such as pleural effusion, sepsis, and respiratory failure since these complications put a child at high risk of death [16]. Moreover, the unrecognised or incorrectly categorised cases may accelerate the transmission in the community and cause the misuse of antibiotics, contributing to antimicrobial resistance [17], [18].

1.2. Challenges in diagnosis and current diagnostic approaches

Clinical manifestations (e.g., tachypnea, chest indrawing, hypoxia) are frequently resorted to in the area with limited resources to prove the diagnosis, yet these signs are not specific and may be similar to other respiratory infections [19]. This points to the relevance of the new and improved diagnostic instruments that may be readily accessible and available to caregivers in order to handle children outcomes and health disparity in the world.

The challenge of an adequate diagnosis of the pneumonia is aggravated with a very wide spectrum of the potential causative agents which vary significantly and depend on the age and immunological history of the child and underlying inoculation status [20]. Whereas, early signs and symptoms, which are stores of pre-diagnostic evidence, viral diseases i.e. respiratory syncytial virus (RSV), parainfluenza, adenovirus, and influenza are the most frequent cause in children less than five [21], [22]. Although bacterial pneumonia is decreasing through conjugate vaccines, *Streptococcus pneumoniae* and *Mycoplasma pneumoniae* continue to take place prominently as the most common bacterial pathogens in vaccinated older children and non-neonates [19]. Pediatric pneumonia disease pathogenesis has been seen to take a cascade of events commencing with microbial invasion, after which there have been inflammatory responses as well as immigration of immune cells in alveoli. The result is the exudation of fluids, decreased oxygen exchange and finally, in more advanced stages, an epithelial necrosis of tissues. Clinical manifestations of these pathophysiological changes manifest themselves in the form of the symptoms of fever, cough, accelerated breathing, chest pain, and other signs, such as exhaustion and feeding and eating problems in some instances and gastrointestinal symptoms. These presentations are so general that considerable amount of time has to be devoted to clinical assessment that is inadequate to make a concrete diagnosis [23]. This is what made imaging modalities very valuable in the adequate determination and evaluation of the level of lung involvement. Chest radiography remains the gold standard for pneumonia diagnosis, providing visual confirmation of characteristic features such as alveolar consolidation, interstitial infiltrates, and pleural effusion [18]. The results give critical pointers to health practitioners in enabling them to determine the degree of the disease and develop relevant treatment protocols. Radiographs are also useful in distinguishing between typical bacterial pneumonia with lobar consolidators, and viral or atypical pneumonias, which may have more diffuse, interstitial patterns in cases of children [24]. Despite its diagnostic value, radiography has limitations. In early or mild infections,

radiographic changes may be subtle or entirely absent. In children with underlying chronic conditions or immune compromise, radiological signs may be atypical [24], [25]. Furthermore, interpretation of pediatric chest X-rays is inherently more challenging due to anatomical and developmental variations in the lungs and thoracic structures, which can influence diagnostic accuracy [26]. Alternative diagnostic approaches such as lung ultrasonography have gained prominence, particularly in resource-constrained settings. Lung ultrasound offers several advantages: it is portable, non-ionizing, and relatively inexpensive, with growing evidence supporting its accuracy in detecting consolidations and pleural effusions. However, its utility depends heavily on the operator's skill and experience, limiting widespread applicability in low-resource or non-specialist settings [27], [28]. Laboratory investigations also play a supporting role in pneumonia diagnosis. Biomarkers such as C-reactive protein and procalcitonin provide indirect evidence of infection and may aid in distinguishing bacterial from viral causes, although they lack sufficient specificity and sensitivity when used alone. Molecular assays for pathogen detection are available in some high-resource settings but remain inaccessible in many regions where the disease burden is greatest [29], [30]. These diagnostic limitations underscore the critical demand for innovative solutions that combine high accuracy with scalability and adaptability across varied healthcare settings.

1.3. Deep learning in medical imaging for pneumonia diagnosis

Deep learning has recently surfaced as a revolutionary approach in medical imaging analysis, demonstrating remarkable potential for faster and more standardized pneumonia classification [31]. The CNN deep learning systems do not require radiologist evaluation like traditional pathways and are able to analyze chest radiographs independently, which allows them to identify minute imaging patterns that can be overlooked in the visual analysis process. This inbuilt ability to learn multi-level feature representations using the pixel data makes CNNs highly appropriate in the analysis of medical images [32], [33]. When trained on large amounts of data, CNN models have shown potential to achieve human expert level diagnostic accuracy and in some cases surpass it, making them a potential alternative to resource-poor settings or a setting without access to specialist radiologists. The technology is particularly relevant in the context of children, as timely diagnoses can prevent rapid furthering of the disease [34]. Automated diagnostic platforms have other benefits as they reduce the discrepancy in diagnosis between different clinics as well as promote reproducibility of diagnosis at different healthcare institutions. Moreover, they also make

it possible to quickly scan the images studies thus triaging urgent cases that require urgent care. However, there are technical challenges involved in clinical adoption of deep learning solutions [35]. X-rays of the chest of children have unique features that contrast with those of adults: not only the size of the thoracic region and the ratio of thoracic sizes and ratios, but also the presence of markers of development, conditions that may complicate the generalizability of the model [20]. These issues are further compounded by the lack of available pediatric imaging data, which are often limited in sample size and contends with unbalanced class distribution and variability of labeling, in the effort needed for training and validation. Critically, model learning behavior is shaped by the choice of CNN architectural elements and most importantly the activation functions [36]. Though ReLU has not become any less popular, some new replacements, including Swish and Mish, have been shown to be more smoothly optimized and faster to use in more difficult classification problems [37]. However, they are yet to accumulate enough research on the usefulness in the classification of pediatric pneumonia, which is an enormous opportunity in this respect, methodologically. Interpretability is another important part of implementing medical AI. It is not necessary not only to have clinical acceptance followed by the algorithmic accuracy, but it is possible to have whether one can explain a decision-making process [38], [39]. By the principle of the visualization of image processing, users can understand the way a model operates with the help of such methods as Grad-CAM by visualizing the areas of the image that are considered to be the most significant to the diagnosis [40]. The refinement of these means of explanation is also necessary since, in this particular way, potential predictions may be valid and applicable in clinical practice. The classification of pneumonia in children introduces even more troubles: the morphological alterations of the lungs at various age levels precondition the alteration of pathological image and age-specific immune response as well as high comorbidity rates (e.g., asthma, congenital defects) is characterized by the overlapping of radiographic patterns that are difficult to identify [41]. It implies that it will be possible to identify such subtle imaging signatures of CNNs solely through sophisticated architectures that are trained on high-quality datasets [42].

The recent progress in deep learning in the diagnosis of pneumonia reveals that CNNs achieve superior classification accuracy, particularly when transfer learning and other architecture-related ideas are employed. Though several studies confirm the CNN effective results in the study on analysis of chest radiographs, they also identify the complexities of the clinical variations problem. Kahwachi and Saed [43] identified InceptionV3 and DenseNet121 as top performers, particularly

when integrated with alternative activation functions. This study underscored the critical role of activation function selection in maximizing CNN efficacy for medical image analysis. Parallel research by Walia et al. [44] implemented a DW-CNN framework incorporating the Swish activation function with VGG-16 transfer learning, achieving 98.5% training accuracy but demonstrating reduced test set performance (79.8%), revealing significant generalization challenges. Reis and Turk [45] introduced COVID-DSNet, a model that achieved 100% accuracy in binary classification tasks for COVID-19 detection. While demonstrating exceptional performance in a controlled setting, the generalizability of such highly accurate models to diverse clinical environments and varied imaging protocols remains a critical area of investigation. Sriporn et al. [46] demonstrated that integrating the Mish activation function and the Nadam optimization algorithm into a DenseNet-121 architecture yielded a 98.97% accuracy in identifying pulmonary lesions. This observation highlights the necessity of advanced activation functions, which can reduce information degradation and speed convergence of the models, such as Mish. Wang et al. [47] added Squeeze-and-Excitation (SE) blocks and Parametric ReLU (PReLU) activation to DenseNet. Their optimized model realised F1-score of 94.3%, which shows effectiveness of multi-scale convolution approaches in identifying delicate pathological differentiations against normal anatomical characteristics shown on radiographic images. The reasoned application of the dilated convolutions in their work was also associated with a better performance of the features extraction with a minimum loss of information which, is specifically valuable in case the analysis of high-resolution medical images is performed because the differences in the structure details can be vast. Transfer learning has been especially helpful when it comes to medical diagnostic, at least according to the results made by Luján-García et al. [48]. They have effectively used this with Xception network to classify pediatrics pneumonia obtaining a high value of AUC 97.0%. A good feature of their work was incorporation of the Grad-CAM approach that would enable them to visualize particular part of the radiographs that played the biggest role towards influencing the decisions made by the model. It did not only offer useful clinical information but also made the AI system a lot more transparent, thereby augmenting confidence in the diagnostics work of the AI system. The study by Khan et al. [49] employed any of three pre-trained models known as EfficientNetB1, NasNetMobile, and MobileNetV2 with multiple forms of classification of chest X-rays. EfficientNetB1, especially utilizing the Swish activation parameter successfully surpassed the rest due to its ability to identify accurate results by classifying COVID-19 and viral pneumonia,

the COVID-19 or lung opacity, and a normal case up to 96.13%. This paper highlights the importance of activation functions selection and careful tuning of the model so as to improve classification performance. Even the activation function selection is a delicate process, as highlighted by Mohammed et al. [50], there is no way to find the single-best variant. The best activation function is to a great extent on the given task and the structure of the neural network. Among the most popular are ReLU, Leaky ReLU, Swish, and Mish, but even newer directions, such as SReLU, CELU, or ISRLU, also were noted by the authors in relation to the enhanced convergence properties and predictive performance of these activations in some complex domains of medical imaging. Addressing the challenge of limited labeled data, especially prevalent during the COVID-19 pandemic, Fahim et al. [51] implemented a semi-supervised learning approach using EfficientNet and Noisy Student Training. Their model, which incorporated the Mish activation function alongside batch normalization and dropout regularization, achieved a 98% AUC for classifying COVID-19, pneumonia, and normal cases from chest X-rays. This innovative work demonstrates the potential of combining modern activation techniques with semi-supervised strategies in data-scarce environments. Ha Pham and Tran [42] explored the benefits of ensemble modeling by combining the outputs of three powerful CNN architectures: InceptionResNetV2, DenseNet201, and VGG16. They created a model based on the ensemble of their existing model, which they trained on 5,848 pediatric chest X-ray images and which obtained an accuracy of above 95% and an F1-score improvement of 3 compared to single models. Such results can be used as very good evidence of the importance of ensemble methods when aiming to improve the reliability of classification, at the expense of greater computational intricacy. Jain et al. [52] comparatively analyzed their custom CNN models to different pre-trained models such as VGG16, VGG19, ResNet50, and InceptionV3 when it came to the implementation of detecting pediatric pneumonia. Their custom models recorded good results with a validation accuracy of 92.3%. Conversely, the pre-trained models had more diverse outcomes with the highest accurateness of 88.4 and the lowest of 70.9. The authors indicate they find that in carefully designed and optimized variants that are customized to a specific medical imaging problem, weight-efficient custom CNNs can match, and in special cases even exceed, large pre-trained models. With additional causes to write about transfer learning, Kaya [53] investigated the use of transfer learning with DenseNet121 applied to early detection of pediatric pneumonia. Their fine-tuned model was able to show a significant classification accuracy of 95.03% and F1-score of 96.03% on a publicly available dataset. This

demonstrates how the model can make fast and reliable diagnoses in clinical settings, citing that this is a result of the synergy between deep learning, and transfer learning in high stakes health care tasks. Panwar et al. [40] have addressed the issue of combining transfer learning with interpretability characteristics by creating their COVID-19 diagnosing deep learning model using X-ray and CT images. Their model attained a high detection rate of 96.55% with the application of Grad-CAM to provide visual explanations and the breaking up of the overfitting problem by introducing early stopping. By combining the effectiveness of expert performance in diagnosis with a capacity to visually interpret the outputs of the model, this method allowed developing a model that increased trust and allowed clinicians to perform the necessary checks on the AI-derived results.

1.4. Research gap and motivation

Although much progress has been achieved it remains that there are still several barriers to bridging the implementation gap between deep learning diagnosis systems developed in research and clinical practice:

1. Limits on the amount of data: CNN requires large, high-quality datasets to train, whereas the imaging of the pediatric pneumonia has specific data collection issues. Ethics, data-sharing policies of the institutions, and inconsistent imaging protocols across care organizations limit the availability of data significantly.
2. Generalizing issues: A lot of models perform very well on certain datasets, and they do not do well when these trained models are used on other images that were obtained under varying settings. Such restricted generalizability makes them not very useful in clinical applications especially in different clinical practice environments with different standards of imaging or patients.
3. Explainability gap: Deep learning models have black boxed decision-making procedures that generate uncertainty in the minds of clinicians. A prominent feature of developing intuitive explanation approaches such as Grad-CAM is their necessity to clinical adoption and genuine human-AI collaboration.

4. Resource intensity: The substantial computational requirements for training and deploying high-performance CNNs pose significant implementation barriers in resource-limited regions where pneumonia burden is highest.

This study proposes a multifaceted solution to these challenges through novel architectural optimizations. The approach focuses on three key innovations: (1) integration of advanced activation functions (Swish/Mish) to preserve feature information, (2) implementation of multi-scale convolutional strategies for enhanced adaptability, and (3) incorporation of interpretability tools to bridge the AI-clinician communication gap. The resulting framework aims to deliver accurate, generalizable, and clinically transparent diagnostic models.

1.5. Research hypotheses

This dissertation makes the following research hypotheses:

H1: CNN architectures employing Swish/Mish activation functions will demonstrate superior classification accuracy compared to ReLU-based models by mitigating gradient information loss during backpropagation.

H2: Multi-scale convolutional architectures will exhibit improved robustness across heterogeneous pediatric chest X-ray datasets by capturing pathological features at varying receptive fields.

H3: Grad-CAM-enhanced visualization will significantly improve model interpretability, as measured by clinician confidence in AI-generated diagnoses during validation studies.

After formulating these research hypotheses, it is essential to highlight the original scientific contributions of this dissertation. These contributions are directly derived from the identified research gaps and are designed to advance both methodological innovation in deep learning and its clinical applicability in pediatric pneumonia diagnosis.

1.6. Original scientific contributions

This dissertation makes the following original scientific contributions:

1. Improvement of convolutional neural network architectures adapted for medical image analysis using Swish and Mish activation functions with the aim of reducing information loss during training.

2. Increasing the adaptability and performance of convolutional neural network architectures based on multi-scale convolutions and atrous convolutions.
3. Proposal for improving diagnostic reliability and more precise determination of important areas in medical images by visualizing activation maps using the Grad-CAM method.

Ova disertacija donosi sljedeće izvorne znanstvene doprinose:

1. Unaprjeđenje arhitektura konvolucijskih neuronskih mreža prilagođenih za analizu medicinskih slika aktivacijskim funkcijama Swish i Mish s ciljem smanjenja gubitka informacija tijekom treniranja.
2. Povećanje prilagodljivosti i performansi arhitektura konvolucijskih neuronskih mreža zasnovan na višeskalnim konvolucijama i konvolucijama s povećanim pomakom.
3. Prijedlog poboljšanja dijagnostičke pouzdanosti i preciznijeg određivanja bitnih područja u medicinskim slikama vizualizacijom aktivacijskih mapa pomoću metode Grad-CAM.

1.7. Contributions papers

The contributions of this dissertation are summarized through three peer-reviewed publications, each addressing a specific dimension of architectural optimization, methodological innovation, and interpretability in deep learning for pediatric pneumonia diagnosis:

CP1: Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays

CP2: Pediatric Pneumonia Recognition Using an Improved DenseNet201 Model with Multi-Scale Convolutions and Mish Activation Function

CP3: Interpretable Deep Learning for Pediatric Pneumonia Diagnosis Through Multi-Phase Feature Learning and Activation Patterns

2. DISCUSSION

2.1. Methodological framework and contributions

The methodological design of this dissertation was guided by the overarching goal of evaluating and advancing deep learning methods for the automated diagnosis of pediatric pneumonia using chest radiographs. Instead of presenting methods in isolation from results, the framework integrates experimental decisions with critical reflection on their contributions to medical imaging and clinical artificial intelligence.

2.1.1. Dataset curation and preprocessing

Quality and representativeness of dataset is the basis of the development of any deep learning-based medical imaging research. In this dissertation, Chest X-ray Images (Pneumonia) repository was utilized [54], including 4273 cases of pneumonia and 1583 controls (healthy people), which consisted of 5856 radiographs. In order to be compared, and to reduce the biases introduced by the disparate acquisition conditions, the images were rescaled to a consistent resolution, usable in convolutional neural network architectures. In further supporting generalizability, stratified random split was used. Training was done on the 80% of the dataset, and 20 percent was used as a validation set, but the proportions of classes in both subsets remained the same. This balance reduced the risk of inflated performance estimates and ensured that evaluation reflected the ability of models to handle unseen cases. Table 2.1 summarizes the resulting distribution.

Table 2.1. Distribution of the dataset.

Category	Total Images	Training Images	Validation Images
Pneumonia	4273	3418	855
Healthy	1583	1266	317
Total	5856	4684	1172

2.1.2. Data augmentation and robustness

A central methodological contribution of this dissertation is the systematic use of data augmentation to enhance robustness. Pediatric radiographs exhibit considerable variability due to patient positioning, movement, and differences in equipment calibration. Without augmentation, deep learning models risk overfitting to training-specific artifacts. To counter this, clinically plausible transformations such as horizontal flipping, random rotation, translation (shift), and zooming were applied. Crucially, augmentation was implemented during training, so that each epoch presented the network with a dynamically varied dataset [55]. This strategy not only improved generalizability but also provided a controlled way to study how augmentation interacts with architectural and activation function choices.

2.1.3. Model construction and transfer learning

Alongside robust data handling, this dissertation incorporated a comparative exploration of CNN architectures to understand how design choices affect performance. Four models were selected: InceptionV3, DenseNet201, InceptionResNetV2, and MobileNetV2. These architectures capture different philosophies of feature extraction:

- Inception networks emphasize multi-scale representation,
- DenseNet promotes feature reuse through dense connectivity,
- ResNet variants highlight the value of residual learning for deep optimization,
- MobileNetV2 prioritizes efficiency, making it suitable for low-resource deployment.

To maximize learning efficiency, all models were initialized with ImageNet weights and subsequently fine-tuned on the pediatric pneumonia dataset. This transfer learning strategy accelerated convergence and improved performance, compensating for the moderate dataset size relative to typical deep learning benchmarks.

2.1.4. Research methodology

Taken together, the research methodology can be conceptualized in three sequential steps:

1. Data curation, preprocessing and augmentation
2. Model construction, transfer learning, and optimization

3. Performance evaluation, metric analysis, and model validation

This structured pipeline is depicted in Figure 2.1, which illustrates the flow from dataset preparation through augmentation, model training, and validation.

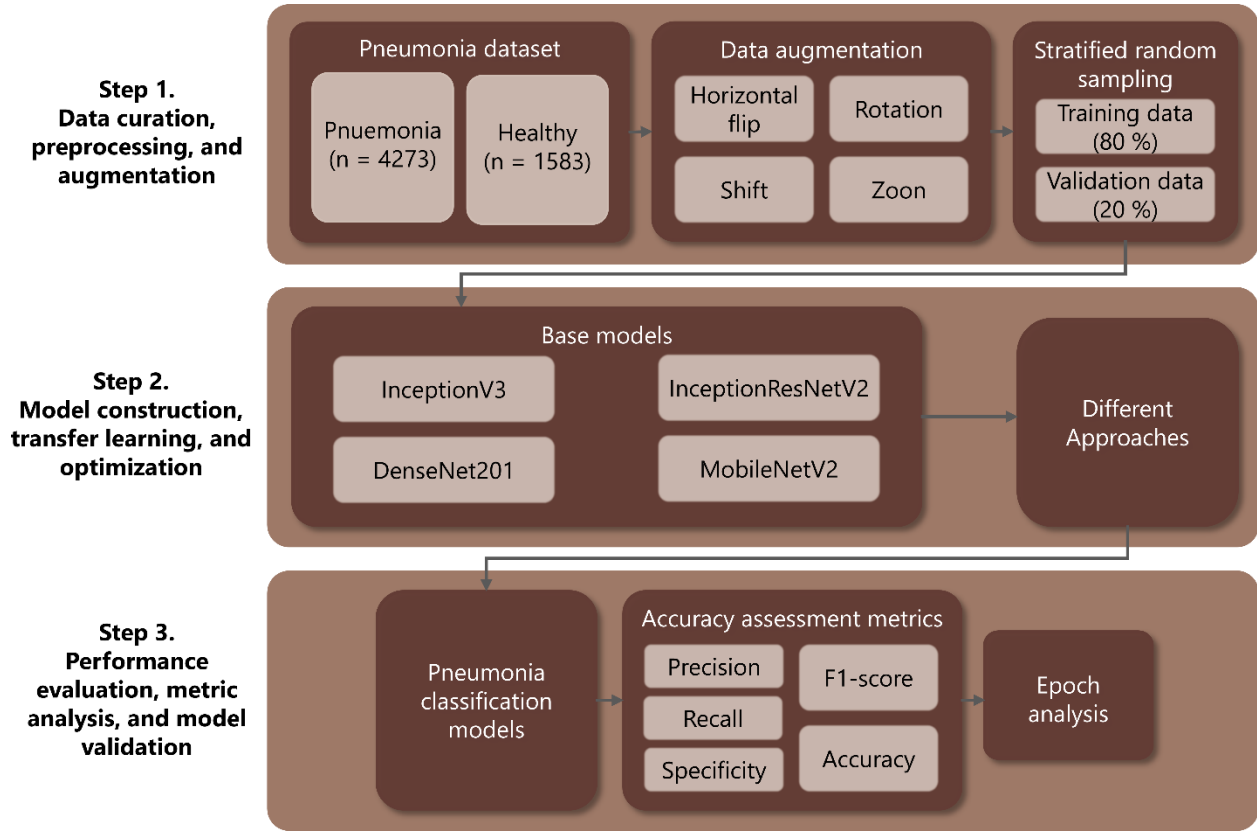


Figure 2.1. Overall workflow of data curation, preprocessing, model construction, and validation.

2.1.5. Validation strategies

The other cornerstone of the methodology is validation robustness. Although simple train-validation splits are widely used in the literature, they are likely to overfit performance and obscure generalizability [48], [56], [57]. To overcome this, stratified validation was always used and the distribution of classes in subsets was always balanced. Besides that, cross-validation experiments using k-fold repeated applying to the data similar to those that are designed to check consistency offered additional evidence.

Performance appraisal was not related to correctness anymore. The entire complex of measures, which comprised precision, recall, F1-score and specificity, was used. The multidimensional view provided a more clinically based evaluation, with error rates in which the missed cases of

pneumonia (false negative) being identified having larger risks as opposed to incorrect positive cases (false positives).

2.1.6. Methodological contributions

The contribution to methodology made by this dissertation can be summed up in three axes:

- Dataset-centric contributions: rigorous preprocessing, balanced stratification, and clinically informed augmentation.
- Model-centric contributions: systematic integration of diverse CNN architectures with transfer learning, enabling comparative evaluation.
- Evaluation-centric contributions: adoption of multidimensional metrics and robust validation strategies, moving beyond simplistic accuracy reporting.

These methodological decisions go beyond technical implementation to facilitate the larger research and clinical objectives. From a computational perspective, the study demonstrates that targeted design decisions, such as augmentation, activation functions, and architectural modules, can have a greater impact than raw model size. This lesson can be applied to other fields of medical imaging where data sets are low and heterogeneity is high. Clinically, the methodology reflects the realities of deployment: limited access to radiologists in resource-constrained regions, variability in imaging conditions, and the need for efficient yet reliable artificial intelligence (AI) tools. By ensuring that models are both robust and interpretable, this work contributes methodological innovations that speak directly to deployment feasibility.

2.2. Validation strategies and experimental robustness

The validation aspect is a key factor that can be used in making sure that deep learning models of medical imaging are not only performative on training data, but also that they can be used to make predictions on unseen clinical cases. The selection of validation strategies in the context of pediatric pneumonia classification, where diagnostic reliability directly translates into patient care, is not a technical issue of merely technical nature but has clinical, ethical, and translational aspects. This section will describe the methodology of the rationale of the validation framework adopted, will elaborate on the description of the evaluation metrics, and will also discuss how the robustness was tested under various experimental conditions.

The task of model validation in medical imaging differs significantly from that in conventional computer vision. Unlike datasets consisting of natural images, radiological datasets often exhibit limited size, class imbalance, and subtle inter-class variability. In addition, the cost of misclassification is asymmetric: failing to classify pneumonia (a false negative) can have far more severe consequences than a false positive, which may only lead to additional but non-invasive follow-up testing.

To address these challenges, this dissertation adopted a validation philosophy grounded in three principles:

- Stratification and balance to ensure equal class representation across training and validation subsets.
- Metric diversity to capture multiple aspects of model performance beyond overall accuracy.
- Robustness testing to evaluate how model performance withstands variation in data, architecture, and optimization procedures.

2.2.1. Stratified validation split

A stratified 80–20 split was implemented, whereby 80% of the dataset was used for training and 20% reserved for validation. Stratification preserved the proportion of pneumonia and healthy cases in each subset, thus avoiding bias in class distribution. This is particularly critical in medical imaging, where disease prevalence is often lower than in general populations and unbalanced splits may lead to misleadingly optimistic performance. While single splits are common, the stratified approach ensured that the model was consistently exposed to a representative sample of both classes during training and evaluation. For exploratory purposes, limited experiments with k-fold cross-validation ($k = 5$) were also conducted. These provided further evidence that the reported results were not artifacts of a particular random partition but reflected consistent model behavior across folds.

2.2.2. Evaluation metrics

To provide a holistic assessment of performance, five complementary metrics were employed: precision, recall, specificity, F1-score, and accuracy. Each metric captures a distinct dimension of

diagnostic utility, and together they provide a clinically meaningful interpretation of model reliability.

The following notation is used:

- TP (True Positives): correctly identified pneumonia cases.
- TN (True Negatives): correctly identified healthy cases.
- FP (False Positives): healthy cases misclassified as pneumonia.
- FN (False Negatives): pneumonia cases misclassified as healthy.

Precision measures the proportion of correctly identified pneumonia cases among all cases predicted as pneumonia. It reflects the model's ability to avoid false alarms.

$$\text{Precision} = \frac{TP}{TP + FP} , \quad (1)$$

High precision, as shown in Equation (1), reduces unnecessary clinical interventions, which is particularly valuable in resource-limited settings where confirmatory diagnostics are costly.

Recall, also known as sensitivity, measures the proportion of true pneumonia cases that the model successfully identifies.

$$\text{Recall} = \frac{TP}{TP + FN} , \quad (2)$$

In clinical terms, high recall, as shown in Equation (2), ensures that few pneumonia cases are missed, which is critical given the life-threatening consequences of delayed diagnosis in children.

Specificity complements recall by quantifying the proportion of true negatives that are correctly classified as healthy, as shown in Equation (3).

$$\text{Specificity} = \frac{TN}{TN + FP} , \quad (3)$$

Together, recall and specificity provide a balanced view of model performance across both positive and negative classes, which is essential in a diagnostic setting where both under- and over-diagnosis have clinical consequences.

The F1 score represents the harmonic mean of precision and recall:

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} , \quad (4)$$

It balances the trade-off between minimizing false positives and false negatives, and is especially useful in cases where class distributions are imbalanced, as shown in Equation (4).

Accuracy reflects the overall proportion of correctly classified cases.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (5)$$

Although accuracy is intuitive, as shown in Equation (5), it can be misleading in unbalanced datasets, which is why it was always interpreted alongside the other four metrics.

2.2.3. Epoch analysis and learning dynamics

In addition to endpoint validation, epoch-by-epoch analysis was conducted to monitor learning dynamics. This analysis provided insights into convergence stability, overfitting tendencies, and the impact of different architectural and activation function choices. For example, early stopping criteria were used when validation loss plateaued or began to diverge from training loss, preventing unnecessary computation and reducing the risk of overfitting [56], [58], [59]. Learning curves were particularly useful in identifying whether augmentation strategies successfully regularized the models and whether architectural modifications led to smoother convergence.

2.2.4. Robustness considerations

Deep learning models strength is a crucial assessment procedure that is significant in determining reliability and clinical relevance of the models. In this dissertation, robustness was discussed under different complementary aspects.

In assessing architectural robustness, in the first instance, the different CNN backbones, including InceptionV3, DenseNet201, InceptionResNetV2, and MobileNetV2 were compared by doing it in a systematic manner. It was through this comparison that it became possible to determine which architectures are capable of generating high predictive accuracy, as well as of generating consistent performance under varying conditions of training. Second, the optimization was robust when the experiments with varying learning rates and varying optimization algorithms (Adam and stochastic gradient descent with momentum) were implemented. These experiments introduced the idea of how convergence stability and the generalization ability are influenced by the optimization strategy in which poor settings were more likely to lead to the instability of performance even across the same architecture. Third, the data strength was investigated by applying the augmentation strategies of varying magnitudes and testing the models on the datasets that were created

deliberately and were just a bit dissimilar to the training distribution. This step was crucial in the assessment of the appropriateness of the models to the real-life situation when the image quality, the demographics of the patients, or the conditions of acquisition may vary.

All these findings served to underscore the fact that, strength is not an accessory or peripheral matter but a primary determinant of clinical viability. A model with high sensitivity and unstable specificity in all validation folds, e.g., can provide inconsistent and ultimately unreliable diagnostic support. On the other hand, the models maintaining a balanced performance in both strength dimensions are much more apt to be integrated into clinical practice, where stability and reproducibility are of the same importance as raw accuracy.

2.3. The impact of activation function selection

In recent years, DNNs have been making stunning progress in the field of medical image processing, and pneumonia identification in a chest radiograph becomes among the most popular tasks in the domain. The models are found to be useful in forming complex patterns, which are ambiguous or hard to perceive in human beings to any extent [35], [46]. Perhaps, the most crucial design choices of DNN construction are activation functions whose core functionality relates to network behavior learning in an effective manner, generalizing and converging [60]. The form of non-linearity that is employed in order to introduce the network is via the application of activation functions; the functions enable the network to map any type of complex functionality as well as model a more complicated correlation in the input data. Their design is directly affected as of the learning dynamics, representational capability, and execution of gradient-based optimistic algorithms [36].

Traditionally, the default selection has been ReLU owing to ease of use, ease of computation as well as efficacy in alleviating vanishing gradient problems observed in the use of earlier operations such as Sigmoid or Tanh. Nevertheless, ReLU has its own detriments, such as the so-called dying ReLU issue, in which the neurons produce zero outputs regardless of the input and basically stops learning. To counter these disadvantages, a series of more complex activation functions were proposed to improve the learning capacity even more, particularly in a high-stakes sphere such as medical diagnosis [36], [61]. Here, the ordinary ReLU activation was replaced with a fixed one that maintained the advantages of ReLU in terms of sparsity, but much easier gradient flow and

sensitivity to significant features of images. This change made the model more accurate in distinguishing between healthy and diseased parts of the lung hence making classification more precise. This modification of the activation function enables the input data to be represented in a more fine-grained way, which is essential to enable the network to recover more informative and discriminative features important in the processing of visually complex medical images [62]. Other models do not just substitute the normal activation functions, but have a mixed (hybrid) or heterogeneous structure, with various neurons or layers using different types of activation. These architectures take advantage of the different learning properties of each of the functions. As an illustration, networks based on a combination of activation functions have shown the capacity to carry out implicit feature ranking. The lower-index neurons are expected to have a higher activation sensitivity in these models and consequently represent a dominant feature in the network at early stages [36]. This structural characteristic makes pruning after training, which are less influential, without much degradation in accuracy. This is not only an efficient way of making the computation but also helps to make the model interpretable, because one may better understand what key features it prioritizes [60], [61]. The other significant breakthrough in the activation functions study is the emergence of Hard-Swish (H-Swish) that is actively applied in the MobileNetV3 structure [31]. H-Swish is a computationally cost-effective replacement of Swish function. This activation is differentiable, continuous and permits better gradient flow and permits the network to recognize less crude nonlinearities than ReLU. Utilizing MobileNetV3, it has also been used in MobileNetV3 where it has increased speed of inference and classification accuracy with its application particularly appealing in edge computing and mobile deployment usage. These are vital in those medical cases when real time analysis can be needed, like in the case of point of care testing diagnostics [31]. The Mish activation is also rather promising, as the last one raised a lot of interest owing to being non monotonic, smooth and self regularising [37]. There is no sudden transition and Mish has been exhibiting better convergence during training. In contrast to ReLU which has the problem of the sharp output threshold and the gradient disappearing on negative inputs, Mish has a non-zero gradient at these points, enabling the network to learn better even when slightly poor input is present [37]. As applied in deep convolutional networks, Mish has contributed to large performance improvements when included as part of the InceptionResNetV2 network. The Mish-activated model exhibited substantially higher accuracy of around 97.61%, making it superior to the models provided by other configurations using ReLU or Swish functions

in experiment. The reason is that lung images contain mixed patterns of intensity and fluid accumulation location, which are frequently very diffuse and complex to describe. Mish has demonstrated proficiency in capturing the non-trivial interactions between pixels in such images.

As shown in Figure 2.2, ReLU thresholds all negative inputs at zero, leading to sparsity but at the cost of potential information loss. Swish introduces a smooth transition that retains small negative values, which supports better gradient flow and richer learning. Mish further refines this behavior with a smooth, non-monotonic shape, enabling self-regularization and the capture of subtle pixel interactions. These properties explain why Mish tends to achieve superior convergence and classification accuracy in medical imaging tasks compared to ReLU and Swish.

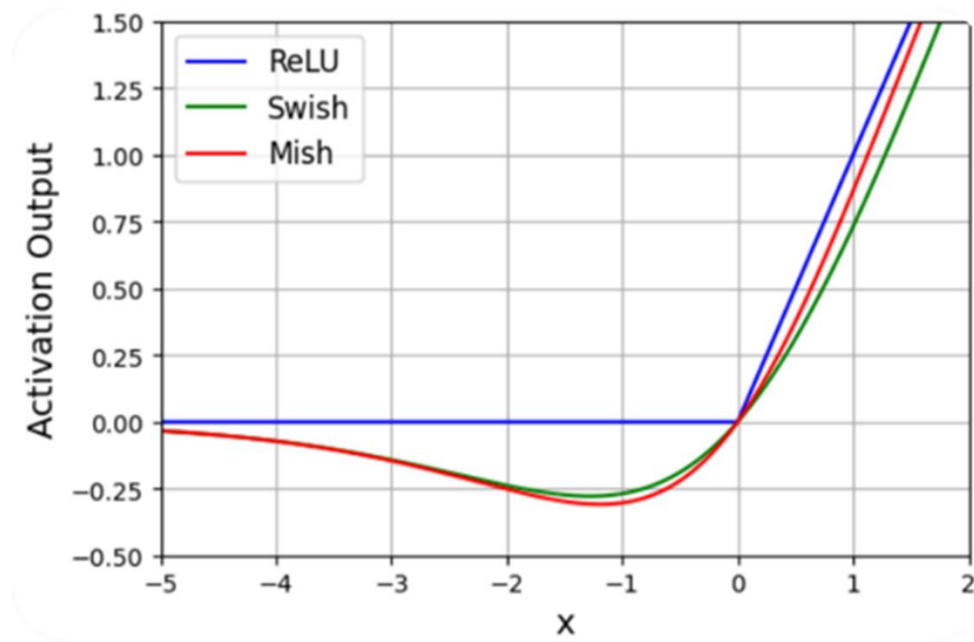


Figure 2.2. Comparative visualization of ReLU, Swish, and Mish activation functions.

A combination of these results, justified in CP1, indicates that the choice of activation functions must be further based not only on their computational effectiveness, but with attention to domain-specific needs of the medical classification task. In medical imaging, data is normally noisy and has slight exceptions and differences among patients. The activation function must also allow the network to be expressive and stable, i.e. capable of learning on small and unbalanced datasets, but sensitive to rare or complex features. Moreover, it is not only the model accuracy that is affected by the selection of the activation functions. These functions like H-Swish and Mish influence the

time of training, convergence rate, model size and energy consumption which are crucial attributes in the context of operations in the real world. As an example, the reduced computational resources or time to apply such an algorithm in clinical practice require faster convergence and less energy consumption to be applicable and scalable to use [63]. Functions with smooth gradient values over larger domains of inputs are more likely to respond well to these methods and produce consistent and stable training. In deep architectures, this synergy is especially significant because, due to gradient degradation, (per layer) effects may compound into one another. Besides, with explainable AI gaining more significance in the fields of medicine, the influence of the activation function on feature attribution and saliency mapping should also be taken into account [38]. The use of activation functions that enable feature representation that is richer and stable enhances the interpretability of models, which can help clinicians make sense of their foundations of a prediction.

2.4. Architectural improvements for feature representation

Neural network architectures have an important impact on defining the efficacy of deep learning networks in task which enables the classification of medical images and specifically the classification of pneumonia [43], [51]. The aim of architectural developments in CNNs is to optimize representational power of models and to trade accuracy and efficiency and complexity of models. Several established approaches to improving the ability of CNNs to extract meaningful features of the chest radiographs have been employed in the last several years, including multi-channel processing, the use of dedicated convolution blocks, and stabilization [32]. The most famous one is probably the InceptionV3 architecture which introduces a number of new features which assist in ensuring efficient and accurate feature extraction. It is founded on a model, which employs the normalization strategy of the inputs of the layers to restrict the internal covariate shift and stabilize the training. It has also applied the convolution factorization; the larger convolutional kernels are divided into smaller convolutional kernels. This reduces the parameters and also increases the computational efficiency of the network [64], [65]. The architecture is also based on different multilayer perceptron (MLP) convolution layers to introduce non-linearity and augment the capability of the model to generalize complex patterns. The Inception module that consists of parallel paths in convolution with different kernels sizes enables the model to process similar features of different scales simultaneously. These advances allow InceptionV3 to reach high

accuracy rates when doing classification because they focus on significant spatial hierarchies and feature abstractions that are vital to identify pneumonia-specific variants of chest X-ray images [32], [34]. In order to complement such complex designs, lightweight architectures like MobileNetV3 have introduced improvements in an effort to improve performance regardless of bound elements [66], [67]. It has depthwise separable convolutions in which conventional convolution operations are separated into two parts, namely depthwise convolution which performs a single filter on each input channel and pointwise convolution that pairs the outputs [67]. In this method the load and the number of parameters is drastically cut down. It is another extension of MobileNetV3, as such it also incorporates ReLU and H-Swish activation functions, strided, and pooling operations into modular blocks. These architecture choices have demonstrated superior classification accuracy with an efficient performance that fits the application on mobile and embedded systems, which makes the model highly applicable in point-of-care diagnosis where the computational requirements are low. Along with more sophisticated networks, simplified ones have been suggested to be enhanced as well [31]. The modified LeNet model attached to the traditional LeNet is a form of the traditional architecture that repeats and concatenates multiple copies of the architecture and adds layers of batch normalization to stabilize learning. It also uses dropout layers in its design, which reduce the possibility of overfitting since a percentage of neurons are dropped randomly as they are trained [62]. These comparatively easy but efficient modifications lead to the better generalization and stability, and claimed precision rates are over the 96% mark in pneumonia categorization tasks [62]. In addition, such networks are attractive to work in real-time screening as they have a shorter running time and require fewer computations to operate [35]. Another architecture trick that has proved to be fruitful has been an introduction of multi-scale and strided convolutions. Such convolutional methods are specially effective to promote feature extraction by varying receptive field and the grain size of amassed spatial data. It is proven that strided convolutions are highly effective in deep architecture like InceptionResNetV2 where it pays attention to localized fine-grained patterns on images [68], [69]. Meanwhile, the multi-scale convolutions are more efficient with the shallower networks such as the DenseNet201 and MobileNetV2 since the networks are capable of detecting more of the semblance in the lung fields. The combination of the two strategies within a given architecture provides the model with an opportunity to model both the global and local characteristics and identify more clearly the complex and spatially variable appearance of pneumonia. To buttress the

operations of models, batch normalization, dropout, and data augmentation will tend to be engaged in structuring the network in such a manner that the functioning becomes stable. In contrast to dropout, batch normalization has the propensity to guarantee the constant flux of the gradient through training and consequently, making a quicker convergence and reducing any vulnerability to the initialisation. The dropout technique increases the power of generalization since the neurons become incompatible with one another, in contrast to data augmentation, which is an artificial increase of the training set that contains variations, i.e., rotation, flipping, and scaling of the data in the training set and therefore, it improves the use of the model to generalize to unseen data [59], [70]. The combination of these building blocks, such as novel and powerful convolution blocks, lightweight design, multi-scale computation and stability layers are helpful in boosting the feature constructing capacity of CNNs altogether. They can make models learn robust, discriminative features based on heterogeneous and usually subtle patterns in the pneumonia radiology images. Such applications are of special concern in clinical settings where precise and prompt diagnosis can make the defining difference, and the visual distinctions between the healthy and pathological may be subtle or obscured by imaging artefacts and inter-patient differences.

Among these strategies, the integration of systematic training procedures and specialized convolutional mechanisms plays a pivotal role. To better illustrate how these architectural principles are operationalized, the following algorithms summarize the main methodological steps undertaken in this research. They encapsulate (Algorithm 1) the end-to-end process of CNN training and evaluation, (Algorithm 2) multi-scale convolutional blocks with Mish activation, and (Algorithm 3) strided convolutional operations with Mish activation.

Algorithm 1: Training and evaluation of CNN for pneumonia recognition

Input: Training dataset directory **train_dir**, Validation dataset directory **validation_dir**, Pretrained base model **CNN_base_model**, Batch size **batch_size**, Number of epochs **epochs**
Output: Trained CNN model, Evaluation metrics (Accuracy, Precision, Recall, F1-score, Specificity), Confusion matrix and classification report

1. **Begin**
2. **Step 1: Initialize training configuration**
 - a. Define callbacks: learning rate reducer on plateau, CSV logger to record training

details

- b. Calculate dataset properties:
 - **nb_train_samples** $\leftarrow$ number of images in train_dir
 - **nb_val_samples** $\leftarrow$ number of images in validation_dir
- c. Define activation functions:
 - $\text{Swish}(x) = x \cdot \text{sigmoid}(x)$
 - $\text{Mish}(x) = x \cdot \tanh(\text{softplus}(x))$

3. Step 2: Configure data augmentation

- a. Initialize ImageDataGenerator with rotation, zoom, width/height shifts, and horizontal flips
- b. Load data generators:
 - **train_generator** $\leftarrow$ training images resized to (224×224)
 - **validation_generator** $\leftarrow$ validation images resized to (224×224)

4. Step 3: Define and compile model

- a. Load base model CNN_base_model (pretrained on ImageNet, include_top = False)
- b. Add custom layers:
 - Convolutional layer with Mish/Swish activation
 - GlobalMaxPooling layer
 - Fully connected layer with sigmoid activation
- c. Compile model with optimizer (Adam), loss (Binary Crossentropy), and metrics (Accuracy, AUC)

5. Step 4: Train the model

- a. Train using **model.fit()** with $\text{steps_per_epoch} = \text{nb_train_samples} / \text{batch_size}$
- b. Validate using $\text{validation_steps} = \text{nb_val_samples} / \text{batch_size}$
- c. Apply callbacks (lr_reducer, csv_logger)

6. Step 5: Evaluate the model

- a. Generate predictions using validation_generator
- b. Compute confusion matrix and classification report
- c. Normalize confusion matrix values

7. **Step 6: Visualize results**
 - a. Plot heatmap of confusion matrix
 - b. Annotate heatmap with class labels (healthy, pneumonia)
8. **Step 7: Save the model**
 - a. Save trained model to disk (e.g., cnn.h5)
9. **End**

Algorithm 2: Multi-scale convolution with Mish activation

Input: Input image, set of convolution kernels $\{\text{Kernel1} \dots \text{KernelN}\}$, corresponding biases $\{\text{Bias1} \dots \text{BiasN}\}$

Output: Combined multi-scale feature map

1. **Begin**
2. **Step 1: Initialize configuration**
 - a. Define convolution kernels at multiple scales (e.g., 3×3 , 5×5 , 7×7)
 - b. Initialize empty list for storing feature maps
3. **Step 2: Apply multi-scale convolutions**

For each scale k in $\{1 \dots N\}$:

 - Perform convolution: $\text{Feature_Mapk} \leftarrow \text{Convolution}(\text{Input_Image}, \text{Kernelk}, \text{Biask})$
 - Apply Mish activation: $\text{Feature_Mapk} \leftarrow \text{Mish}(\text{Feature_Mapk})$
 - Append Feature_Mapk to Feature_Maps
4. **Step 3: Aggregate representations**
 - a. Concatenate all Feature_Maps along the channel dimension
 - b. Normalize aggregated feature maps (Batch Normalization)
 - c. Apply optional dropout for regularization
5. **Step 4: Forward to higher layers**
 - a. Feed $\text{Combined_Feature_Map}$ into subsequent CNN layers for classification
6. **End**

Algorithm 3: Strided convolution with Mish activation

Input: Input image, convolution kernel, stride parameter, bias

Output: Downsampled feature representation

1. **Begin**
2. **Step 1: Initialize parameters**
 - a. Define stride parameter (e.g., stride = 2)
 - b. Select convolution kernel and bias
3. **Step 2: Perform strided convolution**
 - a. Compute feature map using convolution with specified stride
 - b. Apply Mish activation: $\text{Activated_Map} \leftarrow \text{Mish}(\text{Feature_Map})$
4. **Step 3: Post-processing**
 - a. Apply batch normalization to stabilize gradients
 - b. Optionally apply dropout for regularization
5. **Step 4: Integration**
 - a. Pass downsampled representation to subsequent CNN layers
 - b. Combine with multi-scale features if applicable
6. **End**

The inclusion of these algorithms emphasizes that architectural improvements in convolutional neural networks are not restricted to the selection of backbone models, but extend to the systematic integration of convolutional strategies, activation functions, and training protocols. This broader perspective is crucial in medical imaging tasks such as pneumonia recognition, where the discriminative cues in chest radiographs are subtle, variable, and often obscured by anatomical noise. By formalizing these methodological components into algorithmic steps, the design process becomes not only more transparent but also reproducible, which is a prerequisite for deployment in clinical practice.

The training pipeline outlined in Algorithm 1 ensures methodological rigor by standardizing preprocessing, augmentation, and evaluation steps across experiments. Data augmentation plays a particularly central role here, as it allows the model to generalize across variations in patient

posture, imaging quality, and scanner-specific characteristics, which are typical sources of heterogeneity in medical datasets. The use of callbacks such as learning-rate reduction further enhances convergence stability, while the incorporation of multiple evaluation metrics (accuracy, precision, recall, F1-score, specificity) moves the assessment beyond simplistic measures and ensures that models are reliable across different clinical priorities, such as avoiding false negatives in high-risk patients.

Multi-scale convolution (included to the Algorithm 2) is directly related to the problem of variability of scales in medical imaging. Pneumonia in viral pneumonia can manifest as diffuse opacities, small and grained, and in bacterial pneumonia, the pneumonia can manifest as larger, localized pneumonia. All of these various manifestations can be better captured by the network through parallel convolution with various sized kernels. Mish activation is used in this application to improve gradient flow and gradient transitions across features resulting in discriminative, as well as anatomically consistent features maps. This approach is particularly useful, as demonstrated in CP2, where interpretability techniques such as Grad-CAM are applied, where the identified locations are likely to be of clinical interest of the lung.

Strided convolutions, outlined in Algorithm 3, serve a complementary role by enabling efficient downsampling without resorting to pooling operations that may discard valuable spatial detail. Strided convolutions, by implementing downsampling as a part of the convolutional process, maintain the structural continuity and boost the detection of fine-grained pathological patterns by the model particularly in later layers of network models such as InceptionResNetV2. These operations, in combination with Mish activation, give strong representations, which are sensitive to fine pixel-level changes without suffering the problems of vanishing gradients.

2.5. Model interpretability and clinical trust

A predictive performance alone is not sufficient to implement the AI in clinical workflow, namely, diagnostic imaging, but also additional interpretability and transparency along with trust in a clinician [71]. The decision-making paths of deep learning models may be cryptic with the increasing complexity of a specific model, and this attribute of learning models can become a significant barrier in the application of these models to a clinical practice. Interpretability methods fill this gap in some such way as Class Activation Mapping (CAM) and its extensions. They are techniques that offer post hoc visual interpretations of image areas that weight the predictions of

the model the most [26], [40]. Their use in chest radiography model has also shown to be especially useful when mapping outputs of these models to clinically relevant features. A notable method is Grad-CAM, which qualitatively looks at activating the discriminative area of the input images by means of heatmaps created over the convolutional features maps. Grad-CAM has demonstrated the ability to highlight patterns matching radiology results in matching pneumonia diagnosis models [40]. Another example is visual patterns, characteristic of viral pneumonia, which is usually diffuse and spread over both lungs, and sharp and localized in one of the lobes of the lung, in case of bacterial pneumonia [23], [72]. The results that have been obtained by using Grad-CAM visualization mechanisms have made the models attain correspondence to these clinical patterns hence strengthening the models reliability and relevance to diagnosis. The consistency between model attention and clinical intuition reinforces interpretability and boosts the confidence of physicians in model output [73]. Further, more complicated CNN models with concomitant application of Grad-CAM and new activation functions, including Mish, are presented as demonstrating enhancing interpretability and feature richness. Self-regularizing, non-monotonic, Mish activation function has superior effects in terms of flow of gradients and stability of the training process and this facilitates learning when it comes to subtle features [37]. With Grad-CAM, Mish-activated models allow visualizing more continuous and important in a clinical context activation regions. It provides a more anatomically coherent anatomically consistent interpretive result to trace the path of the input images by medical practitioners. This functionality especially comes in handy in the areas of pediatrics and emergency care where quick and effective understanding of a chest radiograph is essential. This interpretive alignment has shown to not only have appeal in theory but has also proven to be verifiable in the practical sense.

To further illustrate these aspects, Figure 2.3 presents Grad-CAM visualizations across four architectures (InceptionV3, InceptionResNetV2, MobileNetV2, and DenseNet201) applied to representative healthy, viral pneumonia, and bacterial pneumonia cases. In this experiment, MobileNetV2 and DenseNet201 were implemented as base models with Mish activation function and multi-scale convolutions, while the Inception networks employed Mish activation with strided convolutions, following the findings of CP3, which reported superior performance under these configurations. Multi-scale convolutional designs, particularly when combined with Mish, were shown to capture hierarchical and fine-grained spatial dependencies, enabling DenseNet201 and MobileNetV2 to highlight subtle pathological regions often overlooked by simpler kernels. On the

other hand, strided convolutions improved computational efficiency and maintained robust feature discrimination in deeper Inception architectures, making them highly suitable for deployment in real-time scenarios. The visualizations highlight how each network emphasizes diagnostically relevant regions: for healthy cases, activations are minimal and concentrated outside pathological areas, whereas in pneumonia cases, particularly viral and bacterial, the networks correctly localize diffuse and lobar patterns of lung involvement. Viral pneumonia activations tend to appear dispersed bilaterally, reflecting its diffuse nature, while bacterial pneumonia heatmaps are sharply localized, aligning with lobar consolidation observed in clinical radiology. Importantly, the Mish activation contributes to smoother and more continuous activation maps, reducing abrupt saliency transitions often seen with ReLU. This property enhances interpretability by providing anatomically coherent attention patterns, which are easier for clinicians to reconcile with radiological findings. Moreover, as shown in the CP3, Mish activation also accelerated convergence and improved classification accuracy, particularly when applied in conjunction with multi-scale convolutional modules, further strengthening its value in medical imaging.

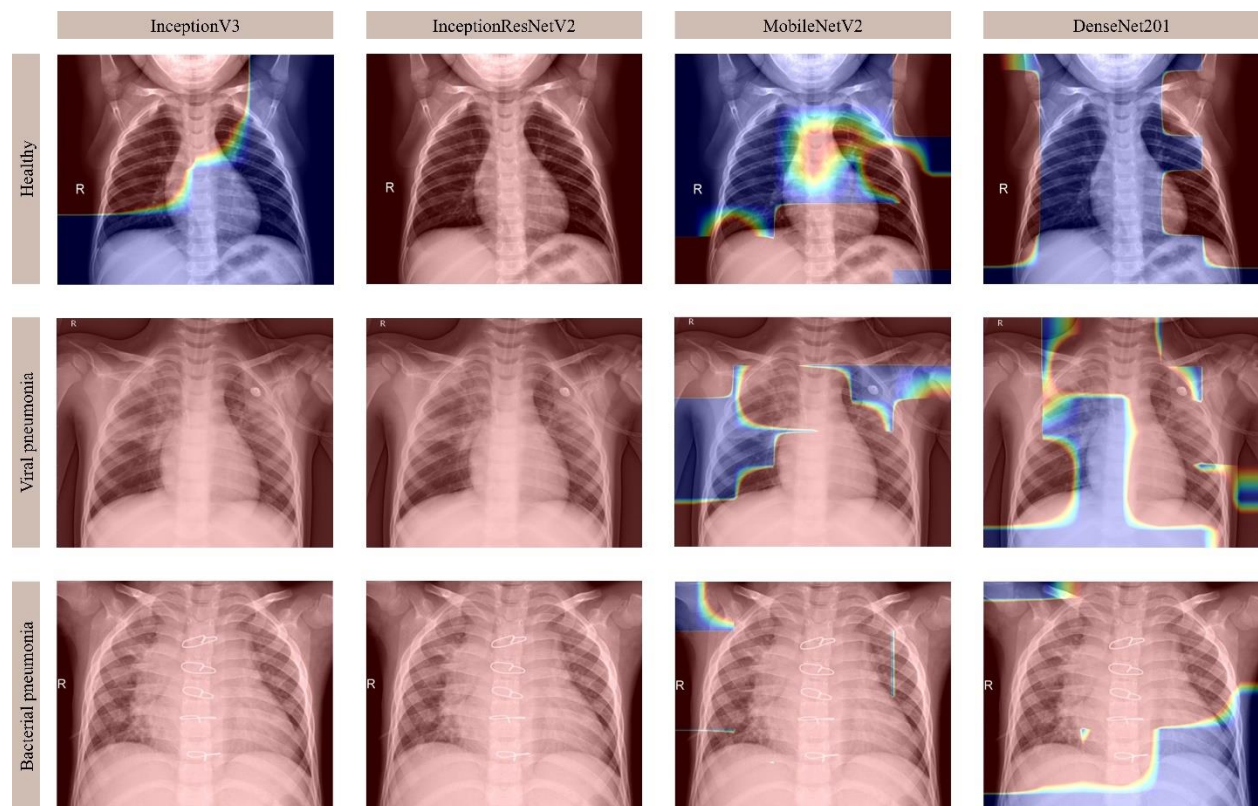


Figure 2.3. Grad-CAM visualizations of healthy, viral pneumonia, and bacterial pneumonia cases across four models. MobileNetV2 and DenseNet201 were trained with Mish activation and multi-scale convolutions, while Inception networks adopted Mish activation with strided convolutions.

Furthermore, comparative studies have revealed that saliency maps identified by the CAM techniques tend to respond to similar regions that interest radiologists. The AI model identified the same portions of abnormality in the lung as radiologists in research concentrating on the diagnosis of *Mycoplasma pneumoniae* [23], [74]. Such overlap on attention regions both generated by machines and the regions that are expected by human beings builds trust and proves confidence on the internal integrity of the model. The clinical specialists will be more inclined to follow the suggestions of the model as a valuable constituent of diagnostics processes when they can observe the AI tool having the appearance of looking at anatomically and pathologically adequate areas [18], [26]. Architecture transparency and its deployment feasibility determine the interpretability and trustworthiness of a model too. Although the diagnosis accuracy can be high with ensemble models and complicated architectures like InceptionResNet or DenseNet, due to their complexity, it may be difficult to implement them in real-time and trace their path of decision-making. Practically, particularly where there are resource constraints, models must be employed, and they must be capable of balancing between performance and interpretability. It is clear that lightweight architectures such as MobileNetV3 built on the expertise of depthwise separable convolutions and small-sized architecture are more computation-efficient. These models are quick and less resource-consuming, thus can be carried out on the bedsides or even in clinics which are located in rural areas. The paper has shown that MobileNetV3 can avoid fidelity to accuracy and interpretability. The model might also effectively give priority to diagnostically significant areas on the chest X-rays based on tight latency budget and going through integrated Grad-cam visualizations. The combination of its compact size and its decipherable nature along with the help of interpretability was friendly to be integrated in mobile health care workstations like the imaging review applications on tablets. Not only is this technically solid but it is also practically viable, especially in situations in which time is of the essence or there is a limitation in infrastructure to support medical facilities. It is also easier to interpret because training methods and elements of architecture are steady and the same. Internal covariate shift is also mitigated by the consideration of batch normalization, and dropout layers offer the mitigation of overfitting, which allows a decision boundary to be more generalizable. Members of these methods equalise the behaviour of

models and reduce the probabilities of unstable or meaningless activations, and this, again, contributes to the need to rely upon design predictions [26]. The steps will assist in creating effective and interpretable AI systems when it is used in combination with post hoc interpretation methods including Grad-CAM. Moreover, jointly trained models with saliency sensitive loss functions or interpretability regularization can be learnt to be consistent with the medical domain. Although the discussed methods are not the norm, it is a necessary medio crepu that will bring model optimization and clinical thinking to the closest [40], [48]. They would prefer to ensure that high predictive accuracy is being achieved at the cost of interpretability but in conjunction with interpretability. The methods can as well be employed to make the model more like the expert decision-making methods by making sure that the model pays attention to the pathologically important areas during the training. The interpretability tools also serve as a critical point of interaction between the human clinicians and the AI systems in regards to human-computer interaction. Clinicians can use them to validate the decisions of a model or dispute them and provide an opportunity to do better based on the trend [18]. As long as clinicians can grasp reasons why a model has chosen a specific decision, they tend to accept such model as a predicative tool that can be looked upon as a collaborator and not as a black box. This is a game-changer and turns a model into a living assistant with the help of which one is able to engage in evidence-based decision-making.

2.6. Comparative performance of deep learning architectures

Machine learning applications to medical images can help greatly in the field of automated pneumonia detection in particular of chest radiographies [46]. The major success factor of these practices is related to the selection of an apt neural network structure which in effect determines the capabilities of any given model to acquire detailed visual features, generalize to various collections and maintain computational complexity [33]. Classic models like VGG16 or AlexNet are classical models that have historical value but competing with newer solutions in a specific medical task may underperform in such situations. As an instance, a designed deep CNN (DCNN) surpassed two popular networks, VGG16 and AlexNet, on a pediatric pneumonia dataset significantly. The custom DCNN obtained an accuracy of 96.09% and a sensitivity of 93.58% regarding specificity of 98.61%, the VGG16 and AlexNet obtained validation accuracy of 92.3% and 93.7 respectively [75]. These findings highlight the fact that architectural simplicity is not sufficient in medical classification tasks, which have to be associated with high sensitivity and

specificity, whereby, model expressiveness and versatility are the pathways. Further, lightweight and shallow architectures, under proper adjustment, are also able to produce competitive results. A smaller architecture based on the LeNet architecture, with concatenated convolutional layers and task-specific additions, an architecture achieved high accuracy values of 96% on a pneumonia chest radiograph dataset [62]. This undercuts the general belief in the overall superiority of deeper networks and indicates that the relation between architecture and task, as well as dataset characteristics (balancedness, size, etc.), are equally essential.

State-of-the-art CNNs ResNet, DenseNet and InceptionResNetV2 have displayed remarkable performance in the classification of pneumonia, in large part to the innovations in terms of feature reuse and graduate stability [47], [61], [76]. Still, architectures with residual connections like ResNet50 can effectively solve the issue of gradient flow remaining between deeper layers, thereby eliminating the problem of the vanishing gradient that majorly occurs in conventional deep networks [77]. Comparatively, the InceptionResNetV2 model with both inception modules and residual connections recorded the highest accuracy of 97.18% after using the Mish activation combination and strided convolutions. InceptionV3 and denseNet201 came next with accuracies of 96.84 and 96.76 respectively. The models took particular strength in classification tasks where we have to distinguish between bacterial and viral pneumonia in the classification task the multi-path based feature extraction approach proves to be very effective in improving the performance on such fine-grained tasks. The practicality of deep CNNs in radiology applications at smaller scale and thus more variable data sets (such as in pediatrics) has also been tested. It is important to note that ResNet50 has become a consistent option as a backbone in that regard. Using the pediatric datasets, a study was performed on ResNet50 structure, where it was shown that after fine-tuning through transfer learning, the model was able to achieve a steady test accuracies of between 82-83%, which outperformed DenseNet121 and EfficientNetV2S [78]. The latter models showed overfitting and reduced generalization even with the more deeper and complex architectures. The higher performance of ResNet-based models in these instances can also be highly explained by the fact that ResNet-based models are designed to feature residual blocks that enable very deep architectures to be trainable because of the maintained identity mappings as well as the potential for a continuous flow of the gradient through these models [78]. These architectural advantages are more significant in the medical areas where the common issues of medical domains include class imbalances and limited data.

Transfer learning has emerged to be a key tactic in medical-imaging use cases in deep learning. The use of weights already learned across large-scale natural visual data (e.g., ImageNet) during initialization of networks gives the networks access to generalized low-level image representations that reduces the data necessary and training costs a great deal to accomplish the medical-relevant task [48], [51]. For instance, ResNeXt50 and WideResNet50, both pretrained and fine-tuned for pneumonia classification, achieved accuracies exceeding 95% within the first few training epochs [79]. Their architectural enhancements, such as grouped convolutions and wide residual blocks, offer additional flexibility and parameter efficiency. Even ShuffleNet, a model optimized for mobile environments, was able to converge to performance levels above 90% after sufficient fine-tuning, despite initially achieving only ~80% accuracy in early epochs [79]. These outcomes affirm that the learning strategy (i.e., from scratch versus transfer learning) often exerts a greater influence on model effectiveness than architectural depth alone.

Although high-performance models like DenseNet201 or InceptionResNetV2 offer impressive accuracy, their computational demands may restrict their usability in low-resource or real-time environments. For clinical deployment in rural areas or mobile health applications, lightweight models such as MobileNet, ShuffleNet, or even customized LeNet derivatives offer a compelling balance between performance, interpretability, and efficiency [62], [79]. Moreover, the integration of interpretability tools such as Grad-CAM or saliency maps further enhances the clinical trustworthiness of these models. Simpler architectures, when combined with transparent feature attribution methods, often provide clinicians with greater confidence in model predictions, especially in high-stakes diagnostic scenarios [26].

In order to put these conclusions into more context, Table 2.2 provides the overview of comparative findings of various CNN models on pneumonia classification in various datasets. The tabulated results are helpful to understand the extent to which architectural design, scale of data, and training strategy have an overall effect on the performance results. In the analysis of the works, it becomes clear that there are strong correlations between performance differences among CNN models and their dependence on the size of datasets, their complexity, and the innovations. For example, Jaganathan et al. [62] report that a modified LeNet-5 architecture reached an accuracy of 96.00% on a dataset of 84,484 radiographs, while Lan et al. [80] observed substantially lower accuracy of 81.00% with DenseNet121 trained on only 578 images. This highlights how limited dataset size

can restrict the generalization capacity of even state-of-the-art models, whereas larger datasets enable simpler architectures to remain competitive with more complex networks.

There has also been strong potential of ensemble approaches. On the experiment with the InceptionResNetV2, DenseNet201, or VGG16, Ha Pham and Tran [42] reached an accuracy of 95.03%, whereas Mabrouk et al. [81] achieved 93.91% with an ensemble of MobileNetV2, DenseNet169, and Vision Transformer. These results imply that ensembles can capitalize on the complementary strengths of individual network designs to address weaknesses arising out of a particular network design. However these strategies are often more computationally intensive and are therefore limited to low resource or in-real-time clinical settings.

Custom CNNs continue to represent a competitive alternative to established architectures. used a tailored DCNN with 96.09 percent accuracy and Stephen et al. [82] and Prakash et al. [83] obtained more than 95 percent accuracy using custom-built designs with medium-scale datasets. Likewise, Sotirov et al. [84] achieved close to 95% accuracy with a simple CNN that was trained on merely 3,000 images, which may further support the fact that even a simple architecture with attentive optimization can achieve high diagnostic accuracy.

Several studies also emphasize the potential of novel architectures and optimization methods. Vrbančić and Podgorelec [85] achieved the highest reported performance in the comparison, with 96.26% accuracy using the SGDRE method. Khan et al. [49] leveraged EfficientNetB1, NasNetMobile, and MobileNetV2 on a dataset of over 21,000 images, reaching 96.13%. These results highlight the scalability and efficiency of modern architectures, which are increasingly attractive for clinical deployment scenarios requiring both robustness and computational economy.

Another determinant is transfer learning. Pretrained on ImageNet and fine-tuned to pneumonia classification, both ResNeXt50 and WideResNet50 both achieved over 95 percent accuracy in the initial training steps [79]. These results further support the significance of the initialisation approaches since initialised weights offer generalisable low-level feature extractors, which significantly lower the data demands and the training expense. In comparison, fully trained models often slow down converging and can overfit more easily especially with small medical datasets.

Table 2.2. Comparative performance of convolutional neural network architectures for pneumonia classification across different datasets.

Reference	Year	CNN Model	No. of Images	Accuracy
Mardianto et al. (2024)	2024	CNN+SVM	6140	0.9200
AlGhamdi (2024)	2024	MobileNetV3	14,000 >	-
Jaganathan et al. (2024)	2024	LeNet-5	84,484	0.9600
Dzhaynakbaev et al. (2024)	2024	VGG16	5228	-
Ha Pham and Tran (2024)	2024	Ensemble (InceptionResNetV2, DenseNet201, VGG16)	5856	0.9503
Mabrouk et al. (2022)	2022	Ensemble (MobileNetV2, DenseNet169, Vision Transformer)	5856	0.9391
Stephen et al. (2019)	2019	Custom CNN	5856	0.9531
Lan et al. (2024)	2024	DenseNet121	578	0.8100
Kaya (2024)	2024	DenseNet121	5856	0.9503
Manickam et al. (2021)	2021	ResNet50	5232	0.9306
Wang et al. (2022)	2022	Custom CNN	5857	0.9280
Vrbančič and Podgorelec (2022)	2022	SGDRE method	5858	0.9626
Yi et al. (2023)	2023	Custom CNN	5856	0.9609
Khan et al. (2022)	2022	EfficientNetB1, NasNetMobile, MobileNetV2	21,165	0.9613
Sotirov et al. (2025)	2025	Custom CNN	3000	0.9493
Prakash et al. (2023)	2023	Custom CNN	7767	0.9615
Kahwachi and Saed (2022)	2022	VGG19, InceptionV3, Resnet50, Inception-ResNetV2, DenseNet121	2482	0.9417
Sharma and Guleria (2023)	2023	VGG16	5856	0.9215
Singh and Tripathi (2022)	2022	Quaternion Convolution neural network	5863	0.9375
CP2			5856	0.9642

Overall, it can be concluded that the findings listed in Table 2 indicate that the most optimal architecture when it comes to pneumonia classification does not exist. Rather, model performance depends on the context and is dictated by the scale of datasets, classes distribution, computational resources, and uniqueness of clinical deployment. Whereas more complex architectures like DenseNet201 and InceptionResNetV2 can demonstrate unparalleled accuracies when using large and equally balanced datasets, less complex models like MobileNet or optimized versions of the LeNet architecture are still more suitable in environments with limited resource availability or diagnostic needs in real-time. The further addition of the interpretability tools, makes the clinical trust even more, which points out that the correspondence between architecture, training strategy, and application domain is the key factor contributing to successful implementation in the medical imaging sphere.

2.7. Convergence dynamics and training stability

The basics of the applied clinical diagnostic systems using deep learning models are efficient and stable training of these models. The convergence speed and stability of training data in tasks involving the processing of medical images such as automated recognition of pneumonia based on radiographs on the chest significantly influence the reliability of the given diagnosis, the model generalizability, and the overall usability of the system [18], [26]. These issues have had a series of solutions to which different approaches have been developed to address these issues including application of transfer learning, more novel architectural design and optimization of hyperparameter tuning. One of the most potent conversion practices of accelerating the rate of convergence is transfer learning. Transfer learning may lead to encapsulating models with low-level and mid-level features that are highly generalizable to visual domains through the initialisation of a model with pretrained weights gained on datasets as large as ImageNet [34]. It was found that the pretrained models, such as ResNeXt50 and WideResNet50, reach over 95% of accuracy, classification with only a few epochs of training [79]. This quicker learning is made possible by such features as edges, corners and textures that are vital in the interpretation of medical images to be reusable. The entirely-trained learning curves of models can be significantly slower and even after extremely long training runs still can be only fractions as accurate by comparison. This gap shows the substantial value of the transfer learning in enhancing the model performance and reducing training durations on clinical setups with limited data.

The next critical element, in stabilizing training, is the architectural structure of the so called neural networks. The original significant departure is that it adopts residual connections as in ResNet architectures. Such connections hasten the propagation of variousiations between layers and, therefore, contribute to the solution of the vanishing gradient problem and train deeper networks without causing a significant decline in performance [79]. The residual units stabilize the backpropagation by providing shortcut routes to the flow of the gradient and have a uniform updating of parameters throughout the network. Such skip connections give strength to a model and simplify training a model. The other effective way of providing stable training is the batch normalization (BN). This makes the BN minimize the covariance shift between the internal variables and the distribution of the layer inputs could be standardized during training in order to be able to increase learning rates. With pneumonia classification models adapting the structure of the LeNet, BN layers inserted between convolutional layers sped up convergence and alleviated problems of gradient explosions [62]. The effects of this normalization were more streamlined patterns in training, and reduced initialization sensitivity that is vital in problems where the medical images are noisy or those of low contrast. Also, the BN led to better regularization, which assisted in avoiding overfitting during the training process.

To improve further on the overfitting and increase generalization, dropout regularization is generally used. At every training cycle, dropout temporarily interpolates a portion of neurons and the values are usually between 0.2 and 0.5, which promotes the network to learn invariant and non-redundant representations [90]. Empirically the use of dropout has been particularly helpful in the case of medical imaging where data are typically small, and subject to overfitting. Based on empirical findings, it has been indicated that dropout, combined with BN, enhances training stability and reliable convergence.

In addition to the more heavy architectures with ResNet, there is an interest in lightweight CNNs, e.g., MobileNetV3 because of the efficiency on device with limited compute capacity. However, there is a tendency of this model to converge more gradually especially within early epochs because of the less capacity of representation. Ahead, the MobileNetV3 underwent evaluation on two datasets with the same architectural settings. This was a controlled setting that permitted the elimination of variability in the dataset and in doing so revealed how disparities in the performance were more probable because of data variability as opposed to architectural variations. Such results

underline the significance of the same experimental conditions when diagnosing performance bottlenecks.

Along with the architectural component, activation functions play an important role in the process of training. New non linear models like Mish and H- Swish have performed better than classic ReLU in its gradient flow and learning aptitude [31]. This has the benefit that these functions are smoother, non-monotone, which can make the optimization more efficient in the deeper networks. As demonstrated in CP1, Mish has already been effectively used with InceptionResNet structures, which allowed achieving improvement in terms of accuracy and acceleration. Selection of the activation function will therefore have a direct effect on how effective the model is at generalization particularly when trained using little data.

Training stability also has an optimization strategy at the core. Other optimizers due to their adaptive nature and the fact that learning dynamics themselves are adjusted based on gradient feedback can be more efficient at converging, including Adam and RMSprop, commonly used with learning rate scheduling [58], [91]. These optimizers, especially, are exceptional at separating noisy and non-stationary gradients, much more likely met in real life, in medical datasets. Moreover, data augmentation and stratified sampling are two more methods that will address convergence by enriching the diversity of data and class imbalance.

The empirical compare shows that the models that use combination of transfer learning, normalization, dropout, and residual connections produce much stronger loss and accuracy using fewer epochs of training. As a case in point, pretrained inceptionresNetV2 and densenet201 models consistently provide results that are above 95% accuracy even as they converge quickly. By comparison, models that have not undergone these enhancements tend to take additional training cycles and are more inconsistent in their results.

2.8. Limitations, generalizability, future directions, and practical implications

Although the current state of innovations in utilizing deep learning models in medical imaging is encouraging, there still exist a number of significant limitations of the research in that field of study, especially regarding the aspects of data diversity and model generalizability [59]. There is worry in that single-source, publicly accessible datasets are used as one of the key factors since they are frequently gathered in a certain clinical setting. Such tight data exclusion constitutes a

possible bias based on the quality of the images, acquisition conditions, and the demographic profile of the patient [92]. Therefore, the models could be ineffective in generalizations when used in other healthcare environments. Domain shift is the phenomenon that shows differences in performance between using a certain model trained over one dataset on new data that are distinct in regards to hardware, image acquisition method, or the population, e.g., age, sex, comorbidities [32], [91]. Experiments that have preceded show that small shifts in the calibration of the X-ray machinery or image protocols would trigger a massive decline in model performance [93], [94]. Thus, the variability in the domains is a critical issue to consider when the design of strong diagnostic systems. Other methods like domain adaptation or style transfer based methods--which seeks to make image distributions similar across domains have been seen to alleviate these effects. The other limitation is the interpretability of deep neural networks. Although methods such as Grad-CAM give a rough idea of how the model works in showing which areas are important in making decisions, it is still inherently a post-hoc explanation [71]. Such heatmaps just imply gradient-based affiliations and certainly not necessarily causal affiliations. This implies that some emphasised area might not be reflective of the real pathological foundation of prediction, but instead some statistical affiliation due to training. Therefore, while Grad-CAM and similar visualizations improve transparency, they cannot fully ensure that the model's reasoning aligns with medical understanding [26], [40]. This limits trust and clinical confidence in automated outputs, especially in high-stakes scenarios such as pediatric diagnostics.

Looking forward, future work should focus on developing models trained on more diverse and multi-institutional datasets, which would improve the reliability and external validity of diagnostic systems. Incorporating data from different geographic, demographic, and technological sources can help mitigate overfitting to specific imaging environments. In addition, improving the internal transparency of models—through architectural innovations or uncertainty quantification—could enhance clinician trust. Such methods may include probabilistic modeling, ensemble predictions, or attention mechanisms designed to offer more interpretable decision pathways.

In summary, while convolutional neural networks have demonstrated high accuracy in classifying pneumonia from chest X-rays, their clinical deployment remains constrained by limitations in dataset diversity and interpretability. It will take more than technical solutions like domain adaptation and more transparent architectures in order to overcome these challenges, but also more

rigorous cross-validation and interdisciplinary collaboration. These underlying issues are the only things that can help such models advance towards more extensive and less risky use in medical diagnostics.

3. CONCLUSIONS

The present doctoral thesis resolves the problems of automatic classification of pediatric pneumonia using deep learning techniques and develops, in some sense, a new field of automatic classification of pediatric pneumonia cases based on chest X-rays in the aspect of model architecture improvements, features extraction smoothness and interpretability. The studies are trying to address basic yet critical issues of medical image research, and these are: to attain greater diagnostic accuracy, to ensure that the model is applicable to the broad range of situations of the imaging phenomenon, and to enhance interpretability of model decision. Another critical part of the dissertation is the integration of new functions of activation, Mish, into CNNs and the development of the architecture with multi-scale and strided convolutions. The study focuses on the explainability methodology as the way of establishing trust within the clinical setting and ensuring that the number of health care professionals who want to use the algorithms will increase in the context of diagnostics.

The most significant discovery of the dissertation is that the selection of the activation functional affects the performance of the deep-learning models in medical imaging significantly. It is empirically confirmed that a replacement of conventional ReLU with less trivial non-monotonic systems such as Mish can provide large improvements, in terms of performance, across most network topologies. The self regulative behavior of Mish in all, accuracy, precision, recall and F1-score, classification tasks achieved higher scores than both ReLU and Swish, which had well-mannered behavior. Advances owe to the capability of Mish to maintain stable gradient propogation, and neural sparsity reduction during training- not a trivial aspect in transfer learning given the relative paucity of data, and the heterogeneity of pediatric chest X-ray data one can find. The proposed Mish-based CNNs reached an accuracy of up to 97.61%, with recall and precision values exceeding 96%, thereby outperforming both traditional ReLU and the more recent Swish activation across all tested backbones.

Besides these optimizations, the dissertation also explores architectural optimizations of CNNs that are termed the multi-scale and strided convolutions. These architectures are meant to enable extraction of features across a variety of receptive fields which are not only covering gross anatomy, but also the finer pathological details. These adaptations gave a stronger

generalization models and higher attenuation to transformation of the form of pneumonia when higher-complex networks, such as DenseNet201, InceptionResNetV2, InceptionV3, are used. Findings verify that multi-scale convolutional layers are representative of the pneumonia associated opacities that can differ extensively in relation to the morphology, size and the spatial distribution. It was this architecture combined with Mish activation which enabled reduced convergence time during training and simultaneously a higher classification accuracy which was less foldsensitive. Comparative evaluations demonstrated that DenseNet201, when augmented with multi-scale, achieved notable gains in sensitivity for both viral and bacterial pneumonia, with overall classification performance surpassing the baseline by a significant margin.

Although the accuracy is still a significant measure of performance when assessing AI-based diagnostic systems, clinical implementation involves models that are potentially transparent and, therefore, comprehensible. To this end, the dissertation made use of the post-hoc explainability tool Grad-CAM. Grad-CAM produces heatmaps, which indicate the areas of the original X-ray images of the reference that the model is using during the prediction. This level of visual interpretability played a critical role in illustrating the diagnostic explanation behind the networks and it was discovered that the models using Mish and multi-scale convolutions had more precise and clinically significant activations in comparison to the models using the traditional elements. Two important benefits of this form of interpretable visualization are that it brings clinicians into the reasoning performed by the model and contributes to increasing diagnostic trust, which is a major requirement before the use of AI in high stakes medical settings.

In aggregate, the findings of this study support three key contributions: the first one is that with new activation functions like Mish, it is possible to reduce information loss during training and vastly improve the performance of the models; the second is that the augmentation of models with multi-scale and dilated convolutional methods can enhance the flexibility and robustness of CNNs trained with variable pediatric imaging data, and the third is that the implementation of Grad-CAM visualization can greatly contribute to the interpretability of model predictions, which can be easily accepted by clinicians and integrated into the diagnostic process.

Nevertheless, some challenges remain. The effective training of deep learning models, especially in medical imaging, such as pediatric pneumonia, is often limited by a lack of numerous, high-quality annotated datasets. Challenges surrounding ethical use, accompanying data sharing policies, and even variations in imaging protocols between hospitals contribute to problems in collecting standardized datasets. In addition, robustness of the models trained is often limited by dataset or imaging condition, raising concerns regarding generalizability in real-world settings. While this research has developed advancements that have improved generalizability specifically through architectural changes, the best architectural approach to develop a trained model that can be integrated into heterogeneous, clinical environments remains an open area of research.

Moreover, the computational requirements of more advanced CNN architectures limit their potential deployment in resource-limited contexts where burden of pneumonia is the highest. While the dissertation assessed efficiency-optimized models, including MobileNetV2, within its comparative analysis, future work should investigate model compression, quantization and edge deployment compared to computers in their respective analysis and implementation. Finally, although Grad-CAM offers a convenient viewing workflow, the key focus should be the creation of more practical, user-friendly and interactive explanation systems that can be more readily integrated into clinical radiology systems.

The implication of this dissertation goes further than the immediate technical findings. Practically, this publication demonstrates that artificial intelligence can be used in ways beyond theory in radiology, and can be a dependable co-pilot to a physician, especially in pediatric cases where timeliness and quality of diagnosis has critical implications. By offering the potential to decrease delays in diagnosis and increase consistency, deep learning systems will become acceptable supportive technology to help achieve better patient outcomes, and be even more valuable in high-burden or resource-constrained domains. Another important issue highlighted in this work is the necessity to rely on clinical professionals and AI systems to collaborate with one another. This requires the creation of interpretability techniques that extend beyond the fixed post-hoc heatmaps in to more interactive evidence, contextualized to clinical practice in a manner that will become embedded in clinical workflows.

On a methodological level, the results can also be relevant to the broader area of deep learning; we have shown that domain-specific construction (of activation functions and architectural modules) can result in improvements beyond just scaling up depth or parameters. It is important to note that this result is applicable not only to pneumonia classification, but also to other fields of medical imaging, where it can be seen that diagnosis can be characterized by comparable complexity and heterogeneity, and can be done with reduced computational expenditure to put into clinical practice.

Future research must both enhance predictive power and consider the federated learning and privacy preserving training approaches to address the challenges to data sharing reported to date across institutions. These strategies would enable collaborative development of strong diagnostic models while maintaining patient confidentiality, thus enabling broader global use of AI-based tools. Finally, interdisciplinary collaboration between computer scientists, radiologists, ethicists, and health policy makers should also be prioritized. Only interdisciplinary collaboration can incorporate algorithmic innovation with clinical significance, regulatory environments, and ethical responsibility.

To conclude, this dissertation highlights the opportunities and challenges of using deep learning systems for diagnosing pneumonia in children. The dissertation makes unambiguous technical contributions around activation functions, architectural improvements, and interpretability, however, it also notes that ethical considerations, practical considerations, and considerations for collaboration between human and AI systems will remain a top priority for use in clinical settings. Collectively, these reflections imply a future where AI is not thought of as a black-box substitute for human expertise but rather a transparent, reliable, and ethically aligned partner to help improve the health of children around the world.

REFERENCES

- [1] D. Kulkarni, X. Wang, E. Sharland, D. Stansfield, H. Campbell, and H. Nair, “The global burden of hospitalisation due to pneumonia caused by *Staphylococcus aureus* in the under-5 years children: A systematic review and meta-analysis,” *EClinicalMedicine*, vol. 44, p. 101267, Feb. 2022, doi: 10.1016/j.eclinm.2021.101267.
- [2] I. Rudan *et al.*, “Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries,” *J. Glob. Health*, vol. 3, no. 1, 2013, Accessed: Sep. 14, 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3700032/>
- [3] H. J. Zar and T. W. Ferkol, “The global burden of respiratory disease-Impact on child health: The Global Burden of Respiratory Disease,” *Pediatr. Pulmonol.*, vol. 49, no. 5, pp. 430–434, May 2014, doi: 10.1002/ppul.23030.
- [4] E. Crame, M. D. Shields, and P. McCrossan, “Paediatric pneumonia: a guide to diagnosis, investigation and treatment,” *Paediatr. Child Health*, vol. 31, no. 6, pp. 250–257, Jun. 2021, doi: 10.1016/j.paed.2021.03.005.
- [5] L. Frigati, L. Greybe, S. Andronikou, E. Eber, S. Sunder B Venkatakrishna, and P. Goussard, “Respiratory infections in low and middle-income countries,” *Paediatr. Respir. Rev.*, vol. 54, pp. 43–51, Jun. 2025, doi: 10.1016/j.prrv.2024.08.002.
- [6] S. M and S. Vaithilingan, “Childhood Pneumonia in Low- and Middle-Income Countries: A Systematic Review of Prevalence, Risk Factors, and Healthcare-Seeking Behaviors,” *Cureus*, vol. 16, no. 4, p. e57636, Apr. 2024, doi: 10.7759/cureus.57636.
- [7] Y. D. Genie *et al.*, “Time to recovery from severe pneumonia and its predictors among pediatric patients admitted in South West Region governmental hospitals, South West Ethiopia: Prospective follow-up study,” *Glob. Pediatr.*, vol. 9, p. 100227, Sep. 2024, doi: 10.1016/j.gped.2024.100227.

- [8] K. L. O'Brien *et al.*, "Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study," *LANCET*, vol. 394, no. 10200, pp. 757–779, Aug. 2019, doi: 10.1016/S0140-6736(19)30721-4.
- [9] C. Onwuchekwa, B. Edem, V. Williams, and E. Oga, "Estimating the impact of pneumococcal conjugate vaccines on childhood pneumonia in sub-Saharan Africa: A systematic review," *F1000Research*, vol. 9, p. 765, 2020, doi: 10.12688/f1000research.25227.2.
- [10] R. Izadnegahdar, A. L. Cohen, K. P. Klugman, and S. A. Qazi, "Childhood pneumonia in developing countries," *Lancet Respir. Med.*, vol. 1, no. 7, pp. 574–584, Sep. 2013, doi: 10.1016/S2213-2600(13)70075-4.
- [11] "Pneumonia in Children Statistics - UNICEF DATA." Accessed: Jul. 22, 2025. [Online]. Available: <https://data.unicef.org/topic/child-health/pneumonia/>
- [12] W. E. Isaac *et al.*, "IN-PATIENT PNEUMONIA BURDEN AND CASE FATALITY RATES IN CHILDREN OVER TWO DECADES IN FEDERAL TEACHING HOSPITAL, GOMBE (FTHG)," *West Afr. J. Med.*, vol. 40, no. 11 Suppl 1, p. S10, Nov. 2023.
- [13] G. R. Patil, "Profile on Risk factors of pneumonia among Under-five age group at a Tertiary care hospital," 2014. Accessed: Jul. 22, 2025. [Online]. Available: https://www.semanticscholar.org/paper/Profile-on-Risk-factors-of-pneumonia-among-age-at-a-Patil/625f79164cccf4bdd14b3239bfe7f92ab77ae786?utm_source=consensus
- [14] J. Hu *et al.*, "Streptococcus pneumoniae and Haemophilus influenzae type b carriage in Chinese children aged 12-18 months in Shanghai, China: a cross-sectional study," *BMC Infect. Dis.*, vol. 16, p. 149, Apr. 2016, doi: 10.1186/s12879-016-1485-3.
- [15] B. Wahl *et al.*, "Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15," *Lancet Glob. Health*, vol. 6, no. 7, pp. e744–e757, Jul. 2018, doi: 10.1016/S2214-109X(18)30247-X.

- [16] S. N. Grief and J. K. Loza, “Guidelines for the Evaluation and Treatment of Pneumonia,” *Prim. Care Clin. Off. Pract.*, vol. 45, no. 3, pp. 485–503, Sep. 2018, doi: 10.1016/j.pop.2018.04.001.
- [17] F. M. de Benedictis, E. Kerem, A. B. Chang, A. A. Colin, H. J. Zar, and A. Bush, “Complicated pneumonia in children,” *The Lancet*, vol. 396, no. 10253, pp. 786–798, Sep. 2020, doi: 10.1016/S0140-6736(20)31550-6.
- [18] N. Fancourt *et al.*, “Standardized Interpretation of Chest Radiographs in Cases of Pediatric Pneumonia From the PERCH Study,” *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, vol. 64, no. suppl_3, pp. S253–S261, Jun. 2017, doi: 10.1093/cid/cix082.
- [19] C. Biagi *et al.*, “Pulmonary and Extrapulmonary Manifestations in Hospitalized Children with Mycoplasma Pneumoniae Infection,” *Microorganisms*, vol. 9, no. 12, p. 2553, Dec. 2021, doi: 10.3390/microorganisms9122553.
- [20] C. R. Jutzeler *et al.*, “Comorbidities, clinical signs and symptoms, laboratory findings, imaging features, treatment strategies, and outcomes in adult and pediatric patients with COVID-19: A systematic review and meta-analysis,” *TRAVEL Med. Infect. Dis.*, vol. 37, p. 101825, Oct. 2020, doi: 10.1016/j.tmaid.2020.101825.
- [21] Z.-M. Chen *et al.*, “Diagnosis and treatment recommendations for pediatric respiratory infection caused by the 2019 novel coronavirus,” *WORLD J. Pediatr.*, vol. 16, no. 3, pp. 240–246, Jun. 2020, doi: 10.1007/s12519-020-00345-5.
- [22] K. H. Lee, A. Gordon, and B. Foxman, “The role of respiratory viruses in the etiology of bacterial pneumonia: An ecological perspective,” *Evol. Med. Public Health*, vol. 2016, no. 1, pp. 95–109, Jan. 2016, doi: 10.1093/emph/eow007.
- [23] M. J. Søndergaard, M. B. Friis, D. S. Hansen, and I. M. Jørgensen, “Clinical manifestations in infants and children with Mycoplasma pneumoniae infection,” *PloS One*, vol. 13, no. 4, p. e0195288, 2018, doi: 10.1371/journal.pone.0195288.

- [24] D. S. Balk *et al.*, “Lung ultrasound compared to chest X-ray for diagnosis of pediatric pneumonia: A meta-analysis,” *Pediatr. Pulmonol.*, vol. 53, no. 8, pp. 1130–1139, Aug. 2018, doi: 10.1002/ppul.24020.
- [25] Z. F. Al Nufaiei and K. M. Alshamrani, “Comparing Ultrasound, Chest X-Ray, and CT Scan for Pneumonia Detection,” *Med. Devices Auckl. NZ*, vol. 18, pp. 149–159, 2025, doi: 10.2147/MDER.S501714.
- [26] A. A. Ali, “Interpretable Deep Learning Framework for COVID-19 Detection: Grad-CAM Integration with Pre-trained CNN Models on Chest X-Ray Images,” *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 12, no. 1, Art. no. 1, Jan. 2025, doi: 10.32628/IJSRSET25121158.
- [27] C. Guitart *et al.*, “Lung ultrasound and procalcitonin, improving antibiotic management and avoiding radiation exposure in pediatric critical patients with bacterial pneumonia: a randomized clinical trial,” *Eur. J. Med. Res.*, vol. 29, no. 1, p. 222, Apr. 2024, doi: 10.1186/s40001-024-01712-y.
- [28] S. Kazi, H. Hernstadt, Y.-N. Abo, H. Graham, M. Palmer, and S. M. Graham, “The utility of chest x-ray and lung ultrasound in the management of infants and children presenting with severe pneumonia in low-and middle-income countries: A pragmatic scoping review,” *J. Glob. Health*, vol. 12, p. 10013, 2022, doi: 10.7189/jogh.12.10013.
- [29] M. A. Pereda *et al.*, “Lung ultrasound for the diagnosis of pneumonia in children: a meta-analysis,” *Pediatrics*, vol. 135, no. 4, pp. 714–722, 2015, Accessed: Sep. 14, 2024. [Online]. Available: <https://publications.aap.org/pediatrics/article-abstract/135/4/714/33632>
- [30] D. Yao, Z. Xu, Y. Lin, and Y. Zhan, “Accurate and intelligent diagnosis of pediatric pneumonia using X-ray images and blood testing data,” *Front. Bioeng. Biotechnol.*, vol. 11, p. 1058888, 2023, Accessed: Sep. 14, 2024. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fbioe.2023.1058888/full>
- [31] A. S. AlGhamdi, “Efficient Deep Learning Approach for the Classification of Pneumonia in Infants from Chest X-Ray Images,” *Trait. SIGNAL*, vol. 41, no. 3, pp. 1245–1262, Jun. 2024, doi: 10.18280/ts.410314.

- [32] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
- [33] N. Dzhaynakbaev, N. Kurmanbekkyzy, A. Baimakhanova, and I. Mussatayeva, “2D-CNN Architecture for Accurate Classification of COVID-19 Related Pneumonia on X-Ray Images,” *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 15, no. 1, Art. no. 1, 33/30 2024, doi: 10.14569/IJACSA.2024.0150191.
- [34] H. Necir, Y. Rebai, and F. Nada, “Novel Transfer Learning Approach for Early Detection of Childhood Pneumonia,” Aug. 23, 2023. doi: 10.21203/rs.3.rs-3259692/v1.
- [35] J. Li and W. Wang, “Deployment and Application of Deep Learning Models under Computational Constraints,” in *2023 IEEE International Conference on Big Data (BigData)*, Dec. 2023, pp. 2529–2533. doi: 10.1109/BigData59044.2023.10386270.
- [36] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, “Activation functions in deep learning: A comprehensive survey and benchmark,” *Neurocomputing*, vol. 503, pp. 92–108, Sep. 2022, doi: 10.1016/j.neucom.2022.06.111.
- [37] D. Misra, “Mish: A Self Regularized Non-Monotonic Activation Function,” presented at the British Machine Vision Conference, 2020. Accessed: Sep. 24, 2024. [Online]. Available: <https://www.semanticscholar.org/paper/Mish%3A-A-Self-Regularized-Non-Monotonic-Activation-Misra/8969d85db0abfc2e2739b752d38280745102eb26>
- [38] M. Owais, H. S. Yoon, T. Mahmood, A. Haider, H. Sultan, and K. R. Park, “Light-weighted ensemble network with multilevel activation visualization for robust diagnosis of COVID19 pneumonia from large-scale chest radiographic database,” *Appl. Soft Comput.*, vol. 108, p. 107490, Sep. 2021, doi: 10.1016/j.asoc.2021.107490.
- [39] M. F. Rahman *et al.*, “Machine-Learning-Enabled Diagnostics with Improved Visualization of Disease Lesions in Chest X-ray Images,” *Diagnostics*, vol. 14, no. 16, Art. no. 16, Jan. 2024, doi: 10.3390/diagnostics14161699.

- [40] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, and V. Singh, “A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images,” *Chaos Solitons Fractals*, vol. 140, p. 110190, Nov. 2020, doi: 10.1016/j.chaos.2020.110190.
- [41] W. G. Kunz, M. Patzig, A. Crispin, R. Stahl, M. F. Reiser, and M. Notohamiprodjo, “The Value of Supine Chest X-Ray in the Diagnosis of Pneumonia in the Basal Lung Zones,” *Acad. Radiol.*, vol. 25, no. 10, pp. 1252–1256, Oct. 2018, doi: 10.1016/j.acra.2018.01.027.
- [42] N. Ha Pham and G. S. Tran, “Apply a CNN-Based Ensemble Model to Chest-X Ray Image-Based Pneumonia Classification,” *J. Adv. Inf. Technol.*, vol. 15, no. 11, pp. 1205–1214, 2024, doi: 10.12720/jait.15.11.1205-1214.
- [43] W. T. Kahwachi and K. A. Saed, “A Novel Architecture of Convolutional Neural Network to Diagnose COVID-19 Disease,” *Math. Stat. Eng. Appl.*, vol. 71, no. 4, Art. no. 4, Oct. 2022, doi: 10.17762/msea.v71i4.1080.
- [44] I. S. Walia, M. Srivastava, D. Kumar, M. Rani, P. Muthreja, and G. Mohadikar, “Pneumonia Detection using Depth-Wise Convolutional Neural Network (DW-CNN),” *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 6, no. 23, Art. no. 23, Sep. 2020, doi: 10.4108/eai.28-5-2020.166290.
- [45] H. C. Reis and V. Turk, “COVID-DSNet: A novel deep convolutional neural network for detection of coronavirus (SARS-CoV-2) cases from CT and Chest X-Ray images,” *Artif. Intell. Med.*, vol. 134, p. 102427, Dec. 2022, doi: 10.1016/j.artmed.2022.102427.
- [46] K. Sriporn, C.-F. Tsai, C.-E. Tsai, and P. Wang, “Analyzing Lung Disease Using Highly Effective Deep Learning Techniques,” *Healthcare*, vol. 8, no. 2, Art. no. 2, Jun. 2020, doi: 10.3390/healthcare8020107.
- [47] K. Wang, P. Jiang, J. Meng, and X. Jiang, “Attention-Based DenseNet for Pneumonia Classification,” *IRBM*, vol. 43, no. 5, pp. 479–485, Oct. 2022, doi: 10.1016/j.irbm.2021.12.004.

- [48] J. E. Luján-García, C. Yáñez-Márquez, Y. Villuendas-Rey, and O. Camacho-Nieto, “A Transfer Learning Method for Pneumonia Classification and Visualization,” *Appl. Sci.*, vol. 10, no. 8, Art. no. 8, Jan. 2020, doi: 10.3390/app10082908.
- [49] E. Khan, M. Z. U. Rehman, F. Ahmed, F. A. Alfouzan, N. M. Alzahrani, and J. Ahmad, “Chest X-ray Classification for the Detection of COVID-19 Using Deep Learning Techniques,” *Sensors*, vol. 22, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/s22031211.
- [50] F. A. Mohammed, K. K. Tune, B. G. Assefa, M. Jett, and S. Muhie, “Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature,” *Mach. Learn. Knowl. Extr.*, vol. 6, no. 1, Art. no. 1, Mar. 2024, doi: 10.3390/make6010033.
- [51] Md. A. I. Fahim, F. Naznin, M. A. Moni, and M. Z. Islam, “Improved Transfer Learning Architecture to Classify Covid-19 Affected Chest X-Rays using Noisy Student Pre-training,” in *2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, Aug. 2021, pp. 1–8. doi: 10.1109/ICIEVicIVPR52578.2021.9564125.
- [52] R. Jain, P. Nagrath, G. Kataria, V. Sirish Kaushik, and D. Jude Hemanth, “Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning,” *Measurement*, vol. 165, p. 108046, Dec. 2020, doi: 10.1016/j.measurement.2020.108046.
- [53] M. Kaya, “Feature fusion-based ensemble CNN learning optimization for automated detection of pediatric pneumonia,” *Biomed. Signal Process. Control*, vol. 87, p. 105472, Jan. 2024, doi: 10.1016/j.bspc.2023.105472.
- [54] D. S. Kermany *et al.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *CELL*, vol. 172, no. 5, pp. 1122–+, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.
- [55] M. H. Davila *et al.*, “Measuring the Impact of Data Augmentation Techniques in Lung Radiograph Classification Using a Fractional Factorial Design: A Covid-19 Case Study,” in

2022 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI), Jul. 2022, pp. 1–6. doi: 10.1109/ColCACI56938.2022.9905303.

- [56] F. Maleki, K. Ovens, R. Gupta, C. Reinhold, A. Spatz, and R. Forghani, “Generalizability of Machine Learning Models: Quantitative Evaluation of Three Methodological Pitfalls,” *Radiol. Artif. Intell.*, vol. 5, no. 1, p. e220028, Jan. 2023, doi: 10.1148/ryai.220028.
- [57] K. Z. Xin, D. Li, and P. H. Yi, “Limited generalizability of deep learning algorithm for pediatric pneumonia classification on external data,” *Emerg. Radiol.*, vol. 29, no. 1, pp. 107–113, Feb. 2022, doi: 10.1007/s10140-021-01954-x.
- [58] M. Subramanian, K. Shanmugavadivel, and P. S. Nandhini, “On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves,” *Neural Comput. Appl.*, vol. 34, no. 16, pp. 13951–13968, Aug. 2022, doi: 10.1007/s00521-022-07246-w.
- [59] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, “Reducing Overfitting and Improving Generalization in Training Convolutional Neural Network (CNN) under Limited Sample Sizes in Image Recognition,” in *2020 - 5th International Conference on Information Technology (InCIT)*, Oct. 2020, pp. 300–305. doi: 10.1109/InCIT50588.2020.9310787.
- [60] G. Bingham and R. Miiikkulainen, “Discovering Parametric Activation Functions,” *Neural Netw.*, vol. 148, pp. 48–65, Apr. 2022, doi: 10.1016/j.neunet.2022.01.001.
- [61] A. A. Alkhoully, A. Mohammed, and H. A. Hefny, “Improving the Performance of Deep Neural Networks Using Two Proposed Activation Functions,” *IEEE Access*, vol. 9, pp. 82249–82271, 2021, doi: 10.1109/ACCESS.2021.3085855.
- [62] D. Jaganathan, S. Balsubramaniam, V. Sureshkumar, and S. Dhanasekaran, “Concatenated Modified LeNet Approach for Classifying Pneumonia Images,” *J. Pers. Med.*, vol. 14, no. 3, p. 328, Mar. 2024, doi: 10.3390/jpm14030328.

- [63] J. Jang, H. Cho, J. Kim, J. Lee, and S. Yang, “Deep neural networks with a set of node-wise varying activation functions,” *Neural Netw.*, vol. 126, pp. 118–131, Jun. 2020, doi: 10.1016/j.neunet.2020.03.004.
- [64] L. R. Baltazar *et al.*, “Artificial intelligence on COVID-19 pneumonia detection using chest xray images,” *PloS One*, vol. 16, no. 10, p. e0257884, 2021, doi: 10.1371/journal.pone.0257884.
- [65] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. de la T. Díez, and I. Ashraf, “Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network,” *Diagnostics*, vol. 12, no. 5, Art. no. 5, May 2022, doi: 10.3390/diagnostics12051280.
- [66] H. Yuan, J. Cheng, Y. Wu, and Z. Zeng, “Low-res MobileNet: An efficient lightweight network for low-resolution image classification in resource-constrained scenarios,” *Multimed. Tools Appl.*, vol. 81, no. 27, pp. 38513–38530, Nov. 2022, doi: 10.1007/s11042-022-13157-8.
- [67] Y. Zou, “Research On Pruning Methods for Mobilenet Convolutional Neural Network,” *Highlights Sci. Eng. Technol.*, vol. 81, pp. 232–236, Jan. 2024, doi: 10.54097/a742e326.
- [68] R. Ayachi, M. Afif, Y. Said, and M. Atri, “Strided Convolution Instead of Max Pooling for Memory Efficiency of Convolutional Neural Networks,” in *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT’18), Vol.1*, M. S. Bouhlef and S. Rovetta, Eds., Cham: Springer International Publishing, 2020, pp. 234–243. doi: 10.1007/978-3-030-21005-2_23.
- [69] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, “Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases,” *Comput. Biol. Med.*, vol. 132, p. 104348, May 2021, doi: 10.1016/j.combiomed.2021.104348.

- [70] D. Zhao, G. Yu, P. Xu, and M. Luo, “Equivalence between dropout and data augmentation: A mathematical check,” *Neural Netw.*, vol. 115, pp. 82–89, Jul. 2019, doi: 10.1016/j.neunet.2019.03.013.
- [71] S. Desai and H. G. Ramaswamy, “Ablation-CAM: Visual Explanations for Deep Convolutional Network via Gradient-free Localization,” presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 983–991. Accessed: Mar. 29, 2025. [Online]. Available: https://openaccess.thecvf.com/content_WACV_2020/html/Desai_Ablation-CAM_Visual_Explanations_for_Deep_Convolutional_Network_via_Gradient-free_Localization_WACV_2020_paper.html
- [72] G. İ. Bayhan *et al.*, “Radiographic findings of adenoviral pneumonia in children,” *Clin. Imaging*, vol. 108, p. 110111, Apr. 2024, doi: 10.1016/j.clinimag.2024.110111.
- [73] R. Fu, “Axiom-based Grad-CAM: Towards Accurate Visualization and Explanation of CNNs”.
- [74] R. C. A. de Groot *et al.*, “Mycoplasma pneumoniae carriage evades induction of protective mucosal antibodies,” *Eur. Respir. J.*, vol. 59, no. 4, p. 2100129, Apr. 2022, doi: 10.1183/13993003.00129-2021.
- [75] R. Yi, L. Tang, Y. Tian, J. Liu, and Z. Wu, “Identification and classification of pneumonia disease using a deep learning-based intelligent computational framework,” *Neural Comput. Appl.*, vol. 35, no. 20, pp. 14473–14486, Jul. 2023, doi: 10.1007/s00521-021-06102-7.
- [76] T. S. Arulananth, S. W. Prakash, R. K. Ayyasamy, V. P. Kavitha, P. G. Kuppusamy, and P. Chinnasamy, “Classification of Paediatric Pneumonia Using Modified DenseNet-121 Deep-Learning Model,” *IEEE Access*, vol. 12, pp. 35716–35727, 2024, doi: 10.1109/ACCESS.2024.3371151.
- [77] D. McNeely-White, J. R. Beveridge, and B. A. Draper, “Inception and ResNet features are (almost) equivalent,” *Cogn. Syst. Res.*, vol. 59, pp. 312–318, Jan. 2020, doi: 10.1016/j.cogsys.2019.10.004.

- [78] I. Putra, N. Dewi, P. Lesmana, I. Suryawan, and P. Putra, "Comparison of ResNet-50 and DenseNet-121 Architectures in Classifying Diabetic Retinopathy," *Indones. J. Data Sci.*, vol. 6, pp. 65–73, Mar. 2025, doi: 10.56705/ijodas.v6i1.232.
- [79] C. Gu and M. Lee, "Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-ray Images," *Bioengineering*, vol. 11, no. 4, Art. no. 4, Apr. 2024, doi: 10.3390/bioengineering11040406.
- [80] X. Lan, Y. Zhang, W. Yuan, F. Shi, and W. Guo, "Image-based deep learning in diagnosing mycoplasma pneumonia on pediatric chest X-rays," *BMC Pediatr.*, vol. 24, no. 1, p. 720, Nov. 2024, doi: 10.1186/s12887-024-05204-0.
- [81] A. Mabrouk, R. P. Díaz Redondo, A. Dahou, M. Abd Elaziz, and M. Kayed, "Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks," *Appl. Sci.*, vol. 12, no. 13, Art. no. 13, Jan. 2022, doi: 10.3390/app12136448.
- [82] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *J. Healthc. Eng.*, vol. 2019, no. 1, p. 4180949, 2019, Accessed: Sep. 14, 2024. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2019/4180949>
- [83] J. A. Prakash *et al.*, "Transfer learning approach for pediatric pneumonia diagnosis using channel attention deep CNN architectures," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106416, Aug. 2023, doi: 10.1016/j.engappai.2023.106416.
- [84] S. Sotirov, D. Orozova, B. Angelov, E. Sotirova, and M. Vylcheva, "Transforming Pediatric Healthcare with Generative AI: A Hybrid CNN Approach for Pneumonia Detection," *Electronics*, vol. 14, no. 9, Art. no. 9, Jan. 2025, doi: 10.3390/electronics14091878.
- [85] G. Vrbančič and V. Podgorelec, "Efficient ensemble for image-based identification of Pneumonia utilizing deep CNN and SGD with warm restarts," *Expert Syst. Appl.*, vol. 187, p. 115834, Jan. 2022, doi: 10.1016/j.eswa.2021.115834.
- [86] M. Mardianto, A. Yoani, S. Soewignjo, I. Putra, and D. A. Dewi, "Classification of Pneumonia from Chest X-ray images using Support Vector Machine and Convolutional

- Neural Network.,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 6, 2024, Accessed: Dec. 14, 2024. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=2158107X&AN=178397316&h=bbkLeiE0dbS1Vg9JIZw8zhmy3sHQvnJncxoXwCUxngOCi8FAExvsMRcYcEzKgefSokIsJ27f2CrQpX9tDvAatg%3D%3D&crl=c>
- [87] A. Manickam, J. Jiang, Y. Zhou, A. Sagar, R. Soundrapandiyar, and R. Samuel, “Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures,” *Measurement*, vol. 184, p. 109953, Jul. 2021, doi: 10.1016/j.measurement.2021.109953.
- [88] S. Sharma and K. Guleria, “A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks,” *Procedia Comput. Sci.*, vol. 218, pp. 357–366, Jan. 2023, doi: 10.1016/j.procs.2023.01.018.
- [89] S. Singh and B. K. Tripathi, “Pneumonia classification using quaternion deep learning,” *Multimed. Tools Appl.*, vol. 81, no. 2, pp. 1743–1764, Jan. 2022, doi: 10.1007/s11042-021-11409-7.
- [90] S. Park and N. Kwak, *Analysis on the Dropout Effect in Convolutional Neural Networks*. 2017, p. 204. doi: 10.1007/978-3-319-54184-6_12.
- [91] C.-L. Hung, C. Hsin, H.-H. Wang, and C. Y. Tang, “Optimization of GPU Memory Usage for Training Deep Neural Networks,” in *Pervasive Systems, Algorithms and Networks*, C. Esposito, J. Hong, and K.-K. R. Choo, Eds., Cham: Springer International Publishing, 2019, pp. 289–293. doi: 10.1007/978-3-030-30143-9_23.
- [92] A. P. De Oliveira and H. F. Tadeu Braga, “Artificial Intelligence: Learning and Limitations,” *WSEAS Trans. Adv. Eng. Educ.*, vol. 17, pp. 80–86, Jul. 2020, doi: 10.37394/232010.2020.17.10.
- [93] H. Liz, M. Sánchez-Montañés, A. Tagarro, S. Domínguez-Rodríguez, R. Dagan, and D. Camacho, “Ensembles of Convolutional Neural Network models for pediatric pneumonia

diagnosis,” *Future Gener. Comput. Syst.*, vol. 122, pp. 220–233, Sep. 2021, doi: 10.1016/j.future.2021.04.007.

- [94] M. Segu, A. Tonioni, and F. Tombari, “Batch normalization embeddings for deep domain generalization,” *Pattern Recognit.*, vol. 135, p. 109115, Mar. 2023, doi: 10.1016/j.patcog.2022.109115.

APPENDICES

Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays

Pediatric Pneumonia Recognition Using an Improved DenseNet201 Model with Multi-Scale Convolutions and Mish Activation Function

Interpretable Deep Learning for Pediatric Pneumonia Diagnosis Through Multi-Phase Feature Learning and Activation Patterns

Article

Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays

Petra Radočaj¹, Dorijan Radočaj^{2,*}  and Goran Martinović³ ¹ Layer d.o.o., Vukovarska cesta 31, 31000 Osijek, Croatia; radocajpetra@gmail.com² Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia³ Faculty of Electrical Engineering, Computer Science and Information Technology, Josip Juraj Strossmayer University of Osijek, Kneza Trpimira 2B, 31000 Osijek, Croatia; goran.martinovic@ferit.hr

* Correspondence: dradocaj@fazos.hr; Tel.: +385-31-554-965

Abstract: Pneumonia remains a significant cause of morbidity and mortality among pediatric patients worldwide. Accurate and timely diagnosis is crucial for effective treatment and improved patient outcomes. Traditionally, pneumonia diagnosis has relied on a combination of clinical evaluation and radiologists' interpretation of chest X-rays. However, this process is time-consuming and prone to inconsistencies in diagnosis. The integration of advanced technologies such as Convolutional Neural Networks (CNNs) into medical diagnostics offers a potential to enhance the accuracy and efficiency. In this study, we conduct a comprehensive evaluation of various activation functions within CNNs for pediatric pneumonia classification using a dataset of 5856 chest X-ray images. The novel Mish activation function was compared with Swish and ReLU, demonstrating superior performance in terms of accuracy, precision, recall, and F1-score in all cases. Notably, InceptionResNetV2 combined with Mish activation function achieved the highest overall performance with an accuracy of 97.61%. Although the dataset used may not fully represent the diversity of real-world clinical cases, this research provides valuable insights into the influence of activation functions on CNN performance in medical image analysis, laying a foundation for future automated pneumonia diagnostic systems.

Keywords: convolutional neural network; pediatric pneumonia; Mish activation function; optimization; deep learning

Academic Editors: Chuan-Ming Liu
and Wei-Shinn Ku

Received: 4 November 2024

Revised: 16 January 2025

Accepted: 23 January 2025

Published: 27 January 2025

Citation: Radočaj, P.; Radočaj, D.; Martinović, G. Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays. *Big Data Cogn. Comput.* **2025**, *9*, 25. <https://doi.org/10.3390/bdcc9020025>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pneumonia remains a significant cause of morbidity and mortality among young children beyond the neonatal period [1,2]. Each year, pneumonia affects approximately 150 million children, predominantly in developing countries [3,4]. Among these cases, an estimated 700,000 to 1 million children under the age of five die annually from pneumonia, making it one of the most significant public health challenges globally [5]. In developed countries, although mortality rates are lower, pneumonia continues to be a primary cause of hospitalization among children [6]. In contrast, the incidence of pneumonia is significantly higher in developing countries with limited healthcare access [7]. The progression of pneumonia relies on the host immunological response and is more frequent in vulnerable individuals, such as children under the age of five.

Although viruses are still the primary cause of pneumonia, the introduction of conjugate vaccines targeting *Streptococcus pneumoniae* and *Haemophilus influenzae* has notably

decreased cases of bacterial pneumonia. Currently, in immunized groups, especially in children older than the neonatal period, *Streptococcus pneumoniae* and *Mycoplasma pneumoniae* remain the most common bacterial pathogens linked to pneumonia. [8–10]. Pneumonia is initiated by the pathogens mentioned, which elicit an immune response, leading to inflammation in the lungs. As a result, the air spaces in the lower respiratory tract become filled with white blood cells, fluid, and cellular debris. This accumulation diminishes lung compliance and increases airway resistance, potentially obstructing smaller airways and causing the collapse of distal air spaces. Consequently, this can lead to air trapping and disrupted ventilation–perfusion relationships. In severe cases of pneumonia, significant infection may cause necrosis of the bronchial or bronchiolar epithelium, as well as damage to the pulmonary parenchyma [11]. Common symptoms of pneumonia in children include cough, fever, rapid breathing, difficulty breathing, chest pain, fatigue, loss of appetite, vomiting, and diarrhea [12,13]. Diagnosing pneumonia typically involves a combination of physical examination, medical history, and laboratory tests [14]. Chest X-rays play a pivotal role in the diagnosis of pneumonia, particularly in pediatric cases, where clinical symptoms can overlap with other respiratory illnesses. The primary function of an X-ray is to provide a detailed view of the lungs, enabling clinicians to identify signs of infection, such as lung consolidation, interstitial inflammation, or pleural effusion [15–17]. In pediatric pneumonia, the typical findings on chest X-rays include areas of increased opacity, which indicate the accumulation of fluid or pus in the alveoli due to the infectious process [18]. However, there are limitations to this diagnostic approach, as chest X-rays may not always be definitive, especially in early-stage infections or in children with underlying health conditions. Complementary diagnostic methods, such as lung ultrasound or laboratory tests, can enhance diagnostic accuracy by providing additional information on the etiology of the disease [14,19,20].

Deep learning has emerged as a powerful approach to the diagnosis of pneumonia, especially through the automated classification of chest X-ray images [21]. Traditional diagnostic methods, such as physical exams, medical history, and X-rays interpreted by clinicians, are time-consuming and subject to human error. Deep learning models, particularly convolutional neural networks (CNNs), offer a significant improvement by automating the process of interpreting X-ray images [22,23]. A fully automated deep-learning pipeline for pneumonia diagnosis offers the advantage of rapidly processing and analyzing large volumes of medical images, enabling efficient and accurate identification of the disease. This efficiency is particularly beneficial in resource-limited settings, such as rural and low-income regions. The automation of pneumonia classification using deep learning enables more timely diagnosis, which is critical in preventing the progression of the disease to more severe stages, especially in children [24]. Additionally, deep learning models reduce the variability in diagnosis that may occur due to subjective interpretations by clinicians. This standardization leads to more consistent results, helping to minimize diagnostic errors. The models can also be fine-tuned to identify subtle patterns and early signs of pneumonia which may not be easily detectable by the human eye [25,26]. Despite a high level of effectiveness in using deep learning for medical imaging, the selection of the optimal activation function remains a challenge that can significantly impact model performance. Firstly, there has been a lack of research focusing on which activation function is optimal in certain tasks, such as pediatric pneumonia diagnosis, especially considering the varying requirements of different models and datasets. This research gap provides an opportunity to consider novel activation functions, such as the Mish function, and evaluate their potential to improve diagnostic accuracy and computational efficiency in pneumonia classification. By addressing this research gap, our study seeks to improve the performance of deep learning models, thereby facilitating more accurate, efficient, and accessible diagnostic solutions for

pediatric pneumonia. To address the limitations of current pneumonia diagnostic methods, this study investigates the application of CNNs, leveraging their capacity for processing extensive medical image datasets and enabling rapid, accurate classification. The central objective is to evaluate the impact of different activation functions on CNN performance, with a specific focus on the novel Mish activation function. Specifically, we conduct a comparative analysis of Mish against Swish and ReLU, two established activation functions in CNN architectures. By determining the optimal activation function, this research aims to improve both the computational efficiency and diagnostic accuracy of CNN models for pneumonia classification. Furthermore, these findings are intended to inform future research and the development of automated diagnostic systems, ultimately facilitating more timely and reliable pediatric pneumonia diagnoses for healthcare providers. The key contributions of this paper are as follows:

- This paper evaluated eight unique CNN architectures that employ Mish and Swish activation functions, providing an innovative approach to pneumonia classification by moving beyond the standard ReLU function.
- Through a comparative analysis of Mish, Swish, and ReLU, this study demonstrated the superior performance of Mish and Swish in CNN models, establishing a basis for further exploration of activation functions in medical imaging.
- The architectures are adaptable to varied clinical settings and hardware constraints, while positioning Mish and Swish as promising alternatives to ReLU, encouraging further research into non-standard activation functions to optimize CNNs in medical applications.

The paper is organized as follows: Section 2 examines some of the relevant studies. Section 3 contains a detailed overview of the models and methodologies employed in this research. Section 4 summarizes the findings of the research. Section 5 contains concluding remarks and future works.

2. Related Works

Despite significant progress in automated pneumonia diagnosis using deep learning, the scalability and robustness of these models in real-world healthcare applications remain areas of ongoing investigation and development. Previous research has demonstrated moderate success in optimizing deep learning methods for the detection of pneumonia from medical images.

Kahwachi and Saed [27] evaluated the performance of six CNN architectures (VGG19, InceptionV3, ResNet50, InceptionResNetV2, and DenseNet121) on a dataset of 2482 chest CT images for COVID-19 classification. Their study explored the impact of different activation functions on model performance, finding that certain alternatives to ReLU could enhance accuracy. Among the models tested, InceptionV3 and DenseNet121 demonstrated superior performance when paired with non-ReLU activation functions. However, their findings suggest that the optimal activation function may vary across different CNN architectures. Walia et al. [28] developed a reliable pneumonia diagnostic model using a Depthwise Convolutional Neural Network (DW-CNN) with Swish activation and transfer learning (VGG16). The proposed model, consisting of 10 convolutional layers and three dense layers, was trained on 5216 augmented radiograph images and tested on 624 images. The model achieved a training accuracy of 98.5%, a testing accuracy of 79.8%, and a validation accuracy of 75%. Reis and Turk [29] designed COVID-DSNet, a novel deep learning architecture to classify chest CT and X-ray images for COVID-19 detection. While ReLU was the primary activation function used in the proposed model, Swish was explored in a separate model. In chest CT, COVID-DSNet achieved 97.60% accuracy in triple classification and 100% in binary classification. For chest X-rays, accuracy was 88.34% in quadruple

classification and 99.45% in binary classification. Khan et al. [30] proposed a deep learning-based method for classifying COVID-19 infections using chest X-rays. Three pre-trained models (EfficientNetB1, NasNetMobile, and MobileNetV2) were fine-tuned and regularized to improve performance. In EfficientNetB1, the Swish activation function was used, while ReLU was employed in MobileNetV2. The EfficientNetB1 model accurately classified four classes (COVID-19, viral pneumonia, lung opacity, and normal) with an accuracy of 96.13%. Sriporn et al. [31] explored deep learning models (MobileNet, Densenet-121, Resnet-50) for computer-aided diagnosis of lung lesions. The Densenet-121 model, combined with Mish activation and Nadam optimization, achieved the best performance. On a held-out test set, the model achieved 98.97% accuracy. They demonstrated the potential of deep learning to assist radiologists in lung lesion detection. Mohammed et al. [32] determined that the activation function in a CNN is crucial for medical image analysis. Appropriate non-linear activation functions can substantially improve network performance, but there is no universally superior activation function; the choice depends on the particular task and network architecture. Commonly employed activation functions include ReLU, Leaky ReLU, Swish, and Mish. Other possibilities include SReLU, ISRLU, CELU, and various modifications of sigmoid and tanh. Fahim et al. [33] utilized chest X-rays to assist in diagnosing COVID-19 during the pandemic, despite the challenge of limited labeled medical images. Using a transfer learning approach with the EfficientNet architecture, the classifier incorporated the Mish activation function, batch normalization, and dropout layers to effectively detect COVID-19, pneumonia, and normal cases. The model, enhanced with semi-supervised Noisy Student Training, achieved a high ROC (AUC) score of 98%.

Previous studies have shown that transfer learning models are beneficial for diagnosing pediatric pneumonia, but there is scope for enhancing their computational efficiency and accuracy. Early detection is crucial to minimize the effects of pediatric pneumonia. This research highlights the significance of evaluating activation functions in convolutional neural networks to improve pediatric pneumonia diagnosis from chest X-rays. By balancing computational complexity and model accuracy, we can achieve more effective disease detection. This paper aims to inform future studies into the potential of Mish and Swish to surpass the long-standing ReLU benchmark and foster further optimization efforts in CNN architectures.

3. Materials and Methods

We evaluate activation functions for pneumonia classification through transfer learning using eight novel CNN architectures, including InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2, alongside the ReLU activation function for objective comparison. Our approach consists of three main steps (Figure 1): (1) preprocessing input data into two classes (pneumonia and healthy) using the open-source Chest X-ray Images (Pneumonia) dataset [34]; (2) testing and comparing ReLU, Swish, and Mish activation functions within each model; and (3) performing pneumonia classification and assessing accuracy. This approach enables an objective evaluation of Mish and Swish activation functions, demonstrating their potential to surpass ReLU and advance CNN design for medical imaging.

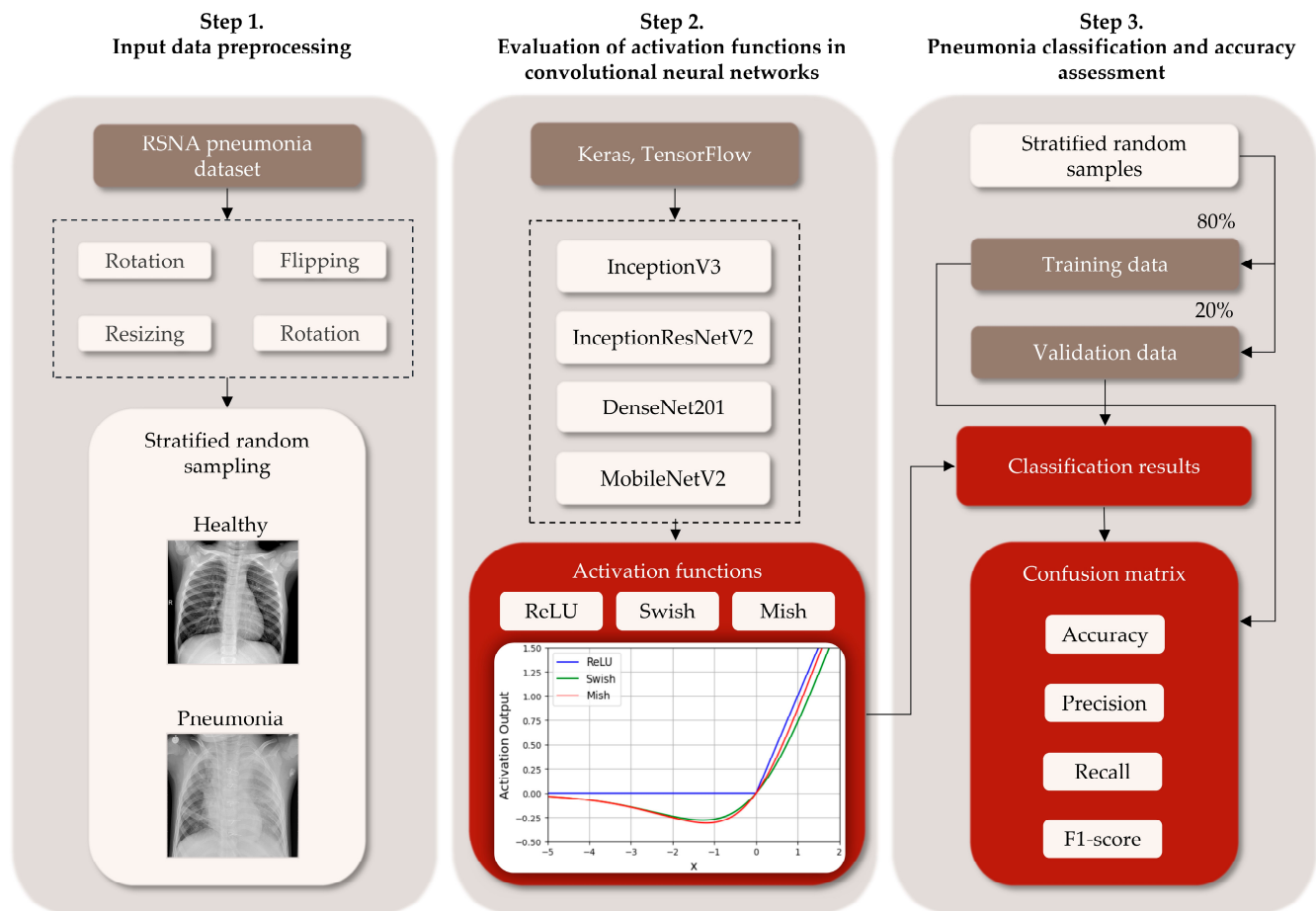


Figure 1. The study workflow for pneumonia classification based on transfer deep learning and evaluated activation functions in the following three steps: (1) input data preprocessing with healthy and pneumonia classes; (2) evaluation of deep learning models with activation functions (ReLU, Swish, Mish); and (3) pneumonia classification and accuracy assessment of evaluated deep learning approaches.

3.1. Data Preprocessing and Experimental Setup

For our experiment, we utilized the open-access Chest X-ray Images (Pneumonia) repository [34], which contains X-ray images collected from pediatric patients aged one to five years at Guangzhou Women and Children’s Medical Center, Guangzhou. These images were captured as part of routine clinical care and are divided into two categories, pneumonia and healthy, with a total of 5856 images [34]. Samples of the images used are shown in Figure 2. To create the training and validation datasets, we applied a stratified random split in an 80:20 ratio, ensuring that the class distribution was maintained in both subsets. The distribution of the dataset is shown in Table 1. To prepare the data for our deep learning model, we resized all images to 224×224 pixels and employed a set of data augmentation techniques during preprocessing. The augmentation techniques, which include rotation, shifting, zooming, and horizontal flipping, were applied following recommendations from previous studies in medical image analysis [35–37]. These techniques ensure that the model learns to generalize well by artificially expanding the training set and introducing variability that the model may encounter in real-world clinical settings.

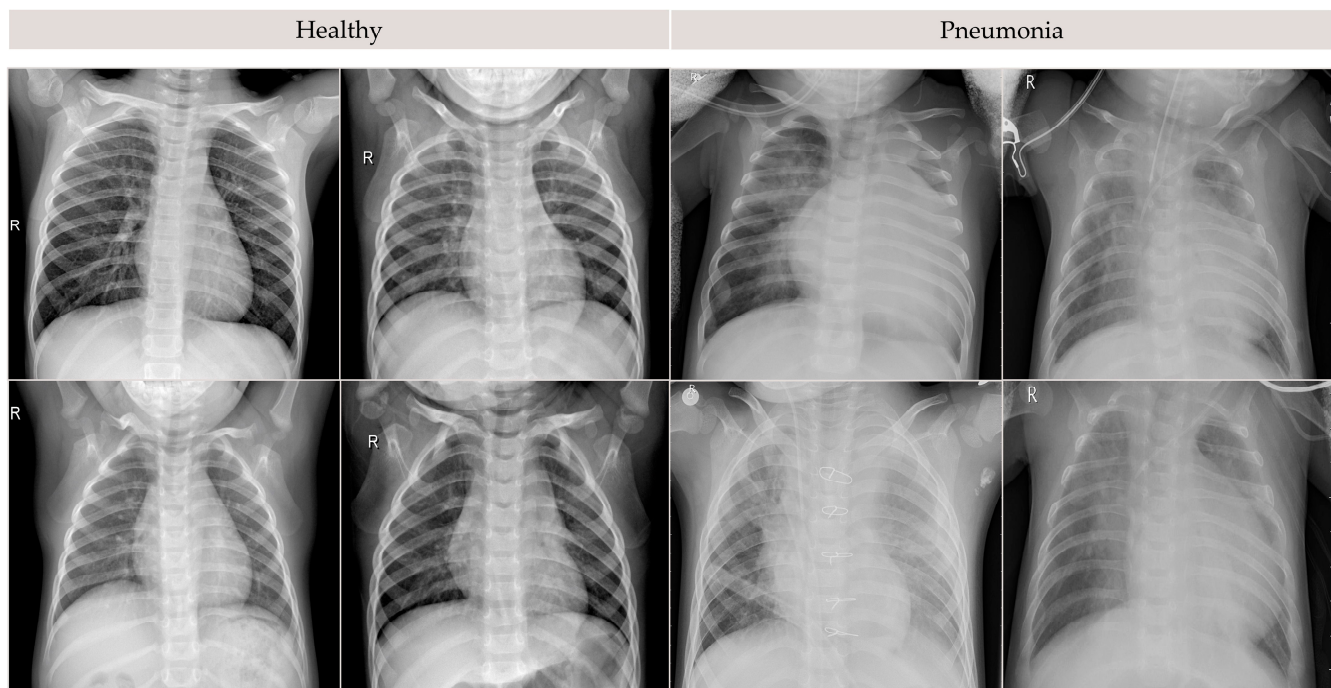


Figure 2. Samples of pneumonia and healthy X-ray images were used in the experiment.

Table 1. Distribution of the dataset.

Category	Total Images	Training Images	Validation Images
Pneumonia	4273	3418	855
Healthy	1583	1266	317
Total	5856	4684	1172

For developing and training the CNN models, we used the Keras-GPU [38] and TensorFlow-GPU [39] platforms in Python. We trained these models on a Google Colab platform, which provided powerful NVIDIA Tesla K80 GPUs with 12 GB of memory. Each model was trained for 20 epochs using a batch size of 32. We used the Adam optimization algorithm to adjust the model's parameters during training. The learning rate, which determines how quickly the model adjusts, was automatically calculated and changed as needed.

3.2. Evaluation of Activation Functions

We examined the importance of selecting activation functions as they introduce non-linearity into neural networks, allowing models to learn complex patterns and relationships in data [40,41]. Without activation functions, neural networks would behave like linear models, limiting their ability to handle complex medical diagnoses. In pediatric pneumonia, where symptoms can vary significantly across patients, a model needs to capture intricate patterns in imaging data and clinical features for accurate classification [42]. Additionally, activation functions enhance the network's capacity to generalize and distinguish between normal and abnormal lung conditions [43,44]. The convolutional layers extract essential features from chest X-rays, and the choice of activation function greatly impacts how these features are transformed and propagated through the network [45]. By comparing different activation functions, we sought to identify the best one for optimizing diagnostic accuracy while ensuring efficient training. Moreover, our experiment aims to enhance the accuracy and robustness of pneumonia detection systems, ultimately supporting earlier interventions

and improved patient outcomes. In this study, we focus on evaluating three activation functions: ReLU, Swish, and Mish. Each of these functions has distinct characteristics that influence how a neural network processes data, particularly in terms of learning complex patterns and improving model performance in pediatric pneumonia recognition.

ReLU is one of the most widely used activation functions in deep learning, primarily due to its simplicity and computational efficiency. It is defined according to Formula (1):

$$\text{ReLU}(x) = \max(0, x) \quad (1)$$

It introduces non-linearity and helps prevent the vanishing gradient problem, which can occur in deeper networks. The key advantage of ReLU is its computational speed and ease of implementation, making it well-suited for large-scale image recognition tasks [46]. In our study, ReLU serves as the baseline activation function due to its widespread use in CNNs.

Swish is an activation function which combines a smooth curve similar to the sigmoid function with a non-linear transformation that retains small negative values, unlike ReLU. It is defined according to Formula (2):

$$\text{Swish}(x) = x \times \sigma(x) = \frac{x}{1 + e^{-x}} \quad (2)$$

Swish, which is introduced in [47], has been shown to outperform ReLU in certain tasks by enabling more complex pattern recognition. In the context of pediatric pneumonia detection, Swish can help the model capture subtle variations in X-ray images, such as slight differences in lung texture that may indicate early stages of pneumonia [28].

Mish is an advanced activation function that has gained attention for its ability to outperform both ReLU and Swish in various computer vision tasks [48,49]. It is defined according to Formula (3):

$$\text{Mish}(x) = x \times \tanh(\text{softplus}(x)) = x \times \tanh(\ln(1 + e^x)) \quad (3)$$

Similar to Swish, Mish which is introduced in [48], is a smooth, non-monotonic activation function that facilitates improved information flow and gradient propagation throughout the network. One of Mish's primary advantages is its capacity for enhanced generalization and more stable training dynamics, especially in deeper networks.

This makes it a promising choice for complex medical diagnoses like pediatric pneumonia, where distinguishing between normal and abnormal lung conditions requires capturing minute details in chest X-ray images.

3.3. Model Architecture for Pneumonia Classification

While numerous studies have investigated CNN-based approaches for pneumonia classification, this paper introduces eight novel architectures leveraging the under-explored Mish and Swish activation functions. The purpose of this design is to objectively assess the efficacy of these non-standard activation functions compared to the widely used ReLU, with the goal of improving general CNN performance. The adaptability of each proposed network to various real-world scenarios enables further fine-tuning in collaboration with medical experts, accommodating specific clinical requirements and hardware constraints.

The model architecture is constructed through a transfer learning approach, utilizing established CNN backbones which are pretrained on ImageNet dataset—InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2. To preserve the generalizable feature extraction capabilities of these pre-trained models, the lower layers are frozen, while the final layers are fine-tuned to specialize in extracting pneumonia-specific features. For effective classification, a custom classification head is introduced. This head includes a

convolutional layer that incorporates the activation function under evaluation—ReLU, Swish, or Mish—enabling a detailed and controlled comparison of the activation functions' impact on model performance. A global max pooling layer follows, which reduces the spatial dimensions by extracting the most salient features, contributing to enhanced robustness. Finally, a fully connected dense layer generates binary predictions, effectively distinguishing between pneumonia and healthy cases. Pseudocode for the methodology used is shown in Algorithm 1.

Algorithm 1: Training and Evaluation of CNN for Pneumonia Recognition

Input: Training dataset directory *train_dir*, Validation dataset directory *validation_dir*, Pretrained base model *CNN_base_model*, Batch size *batch_size*, Number of epochs *epochs*.

Output: Trained CNN model, Evaluation metrics (Accuracy, Precision, Recall, F1-score), Confusion matrix and classification report.

1. **begin**
 2. **Step 1: Initialize Training Configuration**
 3. Define callbacks: *lr_reducer* (reduce learning rate on plateau), *csv_logger* (log training details to CSV);
 4. Calculate dataset properties using *get_nb_files(directory)*: *nb_train_samples* $\leftarrow$ total files in *train_dir*;
 5. *nb_val_samples* $\leftarrow$ total files in *validation_dir*;
 6. Define activation functions: *Swish*(x) = $x \cdot \text{sigmoid}(x)$, *Mish*(x) = $x \cdot \tanh(\text{softplus}(x))$;
 7. **Step 2: Configure Data Augmentation;**
 8. Initialize *ImageDataGenerator* for training with: rotation, zoom, width/height shifts, horizontal flips;
 9. Load data generators: *train_generator* $\leftarrow$ training images resized to (224×224) ;
 10. *validation_generator* $\leftarrow$ validation images resized to (224×224) ;
 11. **Step 3: Define and Compile Model;**
 12. Load base model *CNN_base_model* (pre-trained on *ImageNet*) with *include_top=False*;
 13. Add custom layers: *Convolutional layer* with *Swish/Mish* activation, *GlobalMaxPooling layer*, *Fully connected layer* with *sigmoid* activation;
 14. Compile model with: *Optimizer* = *Adam*, *Loss* = *Binary Crossentropy*, *Metrics* = *Accuracy*, *AUC*;
 15. **Step 4: Train the Model;**
 16. Train using *model.fit()*: *steps_per_epoch* $\leftarrow$ *nb_train_samples/batch_size*;
 17. *validation_steps* $\leftarrow$ *nb_val_samples/batch_size*;
 18. Use callbacks: *lr_reducer*, *csv_logger*;
 19. **Step 5: Evaluate the Model;**
 20. Generate predictions using *validation_generator*;
 21. Compute confusion matrix and classification report;
 22. Normalize confusion matrix values;
 23. **Step 6: Visualize Results;**
 24. Plot heatmap of confusion matrix using *Seaborn*;
 25. Annotate heatmap with class labels (*healthy*, *pneumonia*);
 26. **Step 7: Save the Model;**
 27. Save trained model to disk as *algorithm cnn.h5*;
 28. **end**
-

3.4. Pediatric Pneumonia Accuracy Assessment

To evaluate the accuracy of deep learning models for pneumonia recognition in chest X-ray images, the following metrics were employed: precision, recall, F1-score, and accuracy, according to Formulas (4)–(7). These metrics were calculated using the values from the confusion matrix, providing a comprehensive assessment of the model's performance.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

For pneumonia recognition using chest X-ray images, true negatives are cases where pneumonia is correctly identified. These images often exhibit characteristic patterns such as increased density in the lungs, consolidation, or pleural effusion. False positives are cases where pneumonia is present but not detected. False negatives are cases where pneumonia is incorrectly diagnosed, often due to other conditions or artifacts in the image. True positives are cases where there is no pneumonia present and it is correctly identified as such.

4. Results and Discussion

The accuracy assessment results of pneumonia diagnosis based on four transfer learning models—InceptionV3, InceptionResNetV2, MobileNetV2, and DenseNet201—each evaluated with three activation functions—ReLU, Swish, and Mish—are presented in Table 2. We conducted the experiment using 5856 images to identify pneumonia. Each model's performance was measured by accuracy, F1-score, precision, and recall. DenseNet201 with the Mish activation function achieved the best performance, with an accuracy of 97.53%. Among the activation functions, Mish consistently delivered the highest scores across all metrics for each model, followed closely by Swish, with ReLU generally lagging behind. As demonstrated by Table 2, InceptionV3 achieved 96.33% accuracy with Mish, compared to 91.21% with Swish and 86.95% with ReLU. Similarly, InceptionResNetV2 reached 97.61% accuracy with Mish, while Swish and ReLU yielded 97.44% and 94.37%, respectively. Mish also outperformed other activation functions in MobileNetV2, achieving 92.92% accuracy, compared to 90.53% with Swish and 88.74% with ReLU. DenseNet201 remained the top-performing model, achieving 97.53% accuracy with Mish, compared to 97.27% with Swish and 96.76% with ReLU. Mish consistently provided superior accuracy, F1-scores, precision, and recall, making it the most effective activation function, while Swish remained competitive and ReLU performed the worst across all models.

Some research studies suggest that model performance depends heavily on the choice of activation function [45,50–52], but they do not universally state that Mish is the best activation function for several reasons. Firstly, the effectiveness of an activation function is highly task-specific [52,53]. In this study, we introduced a novel approach by focusing on the application of the Mish activation function in pediatric pneumonia detection. Mish's unique properties, including smooth gradient flow and self-regularization, make it particularly suited for medical image classification tasks requiring deep and complex architectures. Mish has demonstrated superior performance in certain tasks, such as medical image classification, particularly in deep models like DenseNet201 for pneumonia recognition. However, its performance may not generalize across all tasks, datasets, or architectures. For example, simpler or shallow neural networks may not require the advanced properties

of Mish, and activation functions like ReLU or Swish may perform sufficiently well while being computationally more efficient [48]. Secondly, the architecture of the model plays a critical role in determining how well an activation function works [47]. Some models, especially those with complex layers and deep architectures like Inception, might benefit more from the smooth gradient properties of Mish, while other architectures may not see significant gains. Additionally, Mish is computationally more expensive compared to simpler functions like ReLU [54], which can be a critical factor in scenarios where computational resources are limited, or inference speed is a priority. In this study, we demonstrate that Mish's smooth gradient properties enhance the performance of deep architectures like InceptionResNetV2, achieving the highest accuracy. ReLU and Swish have been extensively studied and tested across a wide range of domains, including natural language processing, computer vision, and time-series analysis, providing robust empirical evidence of their reliability [28,54,55]. Moreover, the Swish activation function might offer slightly lower accuracy than Mish in some contexts [48,49], but its lower computational complexity and broader applicability make it a competitive choice.

Table 2. Accuracy assessment of CNNs with evaluated activation functions for pneumonia classification.

Activation Function	Transfer Deep Learning Model	Accuracy	F1-Score	Precision	Recall
ReLU	InceptionV3	0.8695	0.8543	0.8881	0.8695
	InceptionResNetV2	0.9437	0.9415	0.9473	0.9437
	MobileNetV2	0.8874	0.8770	0.9007	0.8874
	DenseNet201	0.9676	0.9673	0.9675	0.9676
Swish	InceptionV3	0.9121	0.9065	0.9197	0.9121
	InceptionResNetV2	0.9744	0.9743	0.9743	0.9744
	MobileNetV2	0.9053	0.9087	0.9247	0.9053
	DenseNet201	0.9727	0.9728	0.9730	0.9727
Mish	InceptionV3	0.9633	0.9630	0.9632	0.9633
	InceptionResNetV2	0.9761	0.9760	0.9760	0.9761
	MobileNetV2	0.9292	0.9259	0.9336	0.9292
	DenseNet201	0.9753	0.9752	0.9752	0.9753

The highest assessment metrics per CNN with evaluated activation functions are in bold.

Although the dataset used was imbalanced, we utilized the F1-score as a robust indicator of test performance, given its harmonic mean of precision and recall. The F1-score results, which support our earlier findings, are detailed in Table 3. InceptionV3 showed the most improvement with the Mish activation function, achieving F1-scores of 93.03% for healthy and 97.51% for pneumonia cases. Swish also yielded strong performance, with F1-scores of 80.82% for healthy and 94.30% for pneumonia cases, while ReLU lagged behind, scoring 68.32% for healthy and 91.78% for pneumonia cases. InceptionResNetV2 demonstrated robust performance across all activation functions. The Mish activation function produced the highest F1-scores of 95.53% for healthy and 98.37% for pneumonia cases, closely followed by Swish with 95.19% for healthy and 98.26% for pneumonia cases. ReLU performed slightly worse, with F1-scores of 88.42% for healthy and 96.28% for pneumonia cases. MobileNetV2 showed more variation between activation functions. Mish again outperformed the other functions, with F1-scores of 85.15% for healthy and 95.35% for pneumonia cases. Swish yielded moderate performance with 84.77% for healthy and 93.13% for pneumonia cases, while ReLU exhibited the lowest scores, particularly for healthy cases with 73.91%, although it performed relatively well for pneumonia cases with 92.82%. DenseNet201, the top-performing model, achieved consistently high F1-scores across all activation functions. Mish resulted in the highest scores, with 95.42% for healthy and 98.31% for pneumonia cases. Swish followed closely with F1-scores of 95.02% for

healthy and 98.12% for pneumonia cases, while ReLU, though effective, produced slightly lower scores of 93.85% for healthy and 97.80% for pneumonia cases.

Table 3. The F1-score values for pneumonia classification per CNN with evaluated activation function.

Activation Function	Transfer Deep Learning Model	Healthy	Pneumonia
ReLU	InceptionV3	0.6832	0.9178
	InceptionResNetV2	0.8842	0.9628
	MobileNetV2	0.7391	0.9282
	DenseNet201	0.9385	0.9780
Swish	InceptionV3	0.8082	0.9430
	InceptionResNetV2	0.9519	0.9826
	MobileNetV2	0.8477	0.9313
	DenseNet201	0.9502	0.9812
Mish	InceptionV3	0.9303	0.9751
	InceptionResNetV2	0.9553	0.9837
	MobileNetV2	0.8515	0.9535
	DenseNet201	0.9542	0.9831

The highest F1-scores for Pneumonia/Healthy class are in bold.

Overall, the results demonstrate that the choice of activation function significantly impacts the performance of each model [50,51]. Among the evaluated activation functions, Mish consistently delivered the best outcomes across all models, surpassing both Swish and ReLU in terms of accuracy, F1-score, precision, and recall. Mish outperforms ReLU and Swish likely due to its unique mathematical properties that balance gradient smoothness and information retention. While both Swish and Mish exhibit non-monotonicity and smooth gradients, Mish's unbounded positive range facilitates more effective information propagation during forward and backward passes, particularly for larger input magnitudes. Furthermore, Mish's sharper curvature near zero introduces implicit regularization, promoting improved feature extraction and enhanced generalization in deep neural networks [48,56]. This makes Mish particularly well-suited for deep learning tasks like pneumonia detection, where accurately capturing subtle differences in medical images is crucial. While Swish also performed well, consistently outperforming ReLU in every model, it fell slightly short of Mish in terms of accuracy and F1-scores. ReLU, though simple and efficient, consistently underperformed compared to the other activation functions, particularly in models requiring deep feature extraction, such as DenseNet201 and InceptionResNetV2. Our study builds upon advancements in deep learning and transfer learning, exploring the efficacy of pretrained models on general-purpose datasets such as ImageNet for medical image analysis. Similarly to the framework of real-world feature transfer learning discussed in research by Jaganathan et al. [57], which achieved an accuracy of 93.6% using DenseNet161, we employ advanced backbone architectures, including InceptionResNetV2 and DenseNet201, to enhance pneumonia classification performance. Furthermore, we incorporate insights from the modified LeNet model, which reported 96% accuracy through a revised ReLU activation function and batch normalization [58]. By integrating novel activation functions, notably Mish, our proposed approach achieves a superior accuracy of 97.61%, significantly surpassing benchmarks established by both DenseNet161 and the modified LeNet. While previous research has emphasized techniques such as grayscale-to-RGB image conversion, batch normalization, and mathematical formalization, our research prioritizes activation function optimization to enhance computational efficiency and robustness for deployment in real-world clinical settings.

The confusion matrix presented in Figure 3 provides a detailed analysis of the classification performance for each model, illustrating the absolute and relative numbers of

correctly and incorrectly classified instances. DenseNet201 consistently outperformed the other models, demonstrating high accuracy and precision in pneumonia diagnosis. InceptionResNetV2 also exhibited strong performance with relatively low false-positive and false-negative rates. Conversely, MobileNetV2 struggled with false positives, particularly in misclassifying healthy cases. InceptionV3 demonstrated moderate performance, with average levels of both false positives and false negatives.

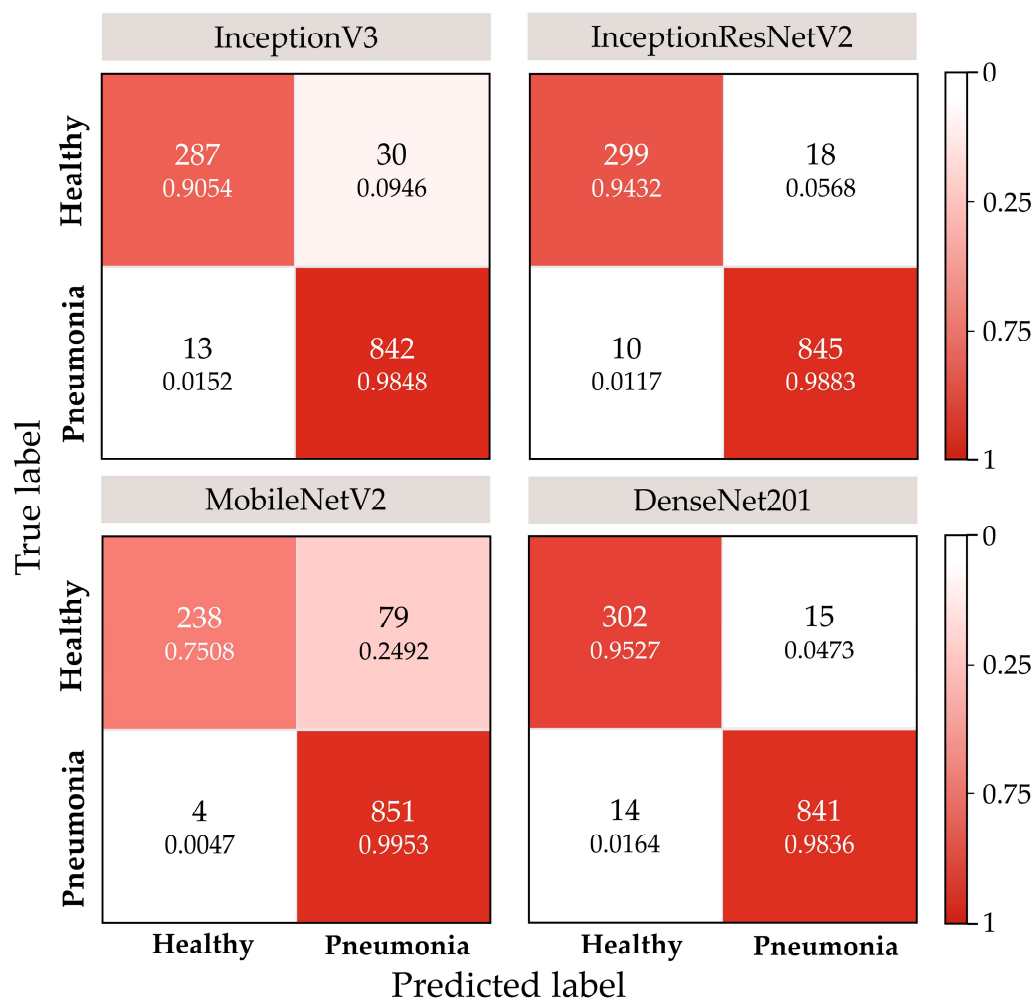


Figure 3. Confusion matrix for CNN with Mish activation function. The upper number in each square represents the classified values, while the lower number indicates the percentage of classified values for each class.

When deployed in a cloud environment, the proposed approach simplifies the implementation of CNN models by eliminating the need for local database storage, significantly reducing storage demands. This study presents eight innovative CNN architectures that leverage non-standard activation functions to enhance the effectiveness of automated pneumonia classification. These architectures are designed for effortless integration into hospital information systems and can be customized to address the specific requirements of various healthcare institutions. By advancing these methodologies, the study promotes the development of reliable deep learning models for pediatric pneumonia recognition, supporting enhanced diagnostic accuracy and timeliness.

In practical applications and clinical settings, this approach presents significant advantages. The cloud-based framework facilitates remote access to diagnostic models, ensuring their availability to healthcare providers, including those in low-resource environments. The flexibility and adaptability of these CNN models enable seamless integration into di-

verse clinical workflows, supporting real-time, data-driven decision-making. By reducing diagnostic inconsistencies and enabling the earlier detection of pneumonia, these models contribute to enhanced healthcare delivery, streamlining diagnostic processes and optimizing patient care. Ultimately, this solution holds the potential to revolutionize pediatric pneumonia management by improving clinical outcomes through faster, more accurate, and widely accessible diagnostic support. While this study demonstrates promising results in the automated diagnosis of pediatric pneumonia using CNNs, several limitations warrant consideration and outline opportunities for future research. A key limitation of this study lies in the dataset used. Although considerable in size, it lacks the diversity and scale necessary to adequately represent the variability observed in real-world clinical contexts. Incorporating a larger and more diverse dataset, including images from a wider range of demographic groups, geographic locations, and clinical settings, is crucial for improving the model's robustness and generalizability. Additionally, Manohar et al. [59] highlight the importance of real-time applications presenting the LSTM-ANN-RSA model which explores adaptive learning mechanisms for real-time data assimilation, while our study focuses on enhancing computational efficiency to enable real-time deployment of diagnostic tools in healthcare environments. The computational complexity of the proposed methods, including training and inference execution times, was not explicitly addressed. These metrics are essential for assessing the practicality of deploying the model in real-time clinical environments.

Future efforts will involve benchmarking the computational requirements and optimizing the model to reduce resource demands while maintaining diagnostic accuracy. Additionally, future research will focus on using methods from explainable AI to analyze the features learned by the model with Mish activation function. This will include examining the lower, unfrozen layers during transfer learning to gain deeper insights into how Mish influences feature extraction and representation. Finally, future research will prioritize real-life testing and integration into clinical workflows to evaluate the model's practical utility and scalability, ensuring its effectiveness in diverse healthcare settings.

5. Conclusions and Future Work

In this study, we demonstrated how CNNs, particularly those utilizing the novel Mish activation function, can significantly enhance pediatric pneumonia classification. By leveraging innovative activation functions, we improved diagnostic accuracy, reduced variability in interpretation, and supported clinical decision-making. Our evaluation of pre-trained CNN architectures, such as InceptionResNetV2, revealed that the Mish activation function outperformed traditional functions like ReLU and Swish, achieving an accuracy of 97.61%. Mish's distinctive properties, such as its ability to maintain smoother gradient flow, self-regularization effects, and non-monotonic behavior, played a pivotal role in enhancing model stability and improving generalization to unseen data. Additionally, Mish's unbounded positive range and sharper curvature near zero allowed for better information retention and feature extraction, particularly in complex datasets like medical images. These attributes collectively reduced the risk of overfitting and contributed to superior performance, establishing Mish as a highly effective activation function for medical imaging tasks. Future research on pediatric pneumonia using deep learning models should focus on multiple key aspects to improve diagnostic precision and efficiency. To optimize model performance, efforts should aim to expand the dataset and ensure a more equitable distribution of healthy and pneumonia examples. Additionally, comprehensive assessments of activation functions, such as Mish, across varied datasets are vital to determine the most effective functions for pneumonia detection in chest X-rays. Furthermore, enhancing computational efficiency is essential for facilitating real-time applications in clin-

ical environments. Another important direction for future research involves investigating the interactions between activation functions and various model architectures. A deeper understanding of these dynamics could facilitate the optimization of neural networks specifically designed for pediatric pneumonia detection, potentially leading to more robust and reliable models as well as improved diagnostic performance. By addressing these priorities, future research can accelerate the development of automated diagnostic tools, equipping healthcare providers with reliable and efficient systems for early and accurate pneumonia diagnosis.

Author Contributions: Conceptualization, P.R.; methodology, P.R.; software, P.R.; validation, D.R. and G.M.; formal analysis, P.R.; investigation, P.R.; resources, P.R.; data curation, P.R.; writing—original draft preparation, P.R.; writing—review and editing, P.R., D.R. and G.M.; visualization, P.R.; supervision, G.M.; project administration, G.M.; funding acquisition, D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The code developed in this study is available on request from the corresponding author. The open access Chest X-ray Images (Pneumonia) repository containing images collected from pediatric patients is divided into two categories (Pneumonia/Healthy) and is available at <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia> (accessed on 24 October 2024).

Conflicts of Interest: The authors declare no competing interests. The author Petra Radočaj was employed by the company Layer d.o.o. There is no conflict of interest between any of the authors and the company Layer d.o.o.

References

1. Zar, H.J.; Ferkol, T.W. The Global Burden of Respiratory Disease-Impact on Child Health: The Global Burden of Respiratory Disease. *Pediatr. Pulmonol.* **2014**, *49*, 430–434. [CrossRef] [PubMed]
2. Rudan, I.; O'Brien, K.L.; Nair, H.; Liu, L.; Theodoratou, E.; Qazi, S.; Lukšić, I.; Walker, C.L.F.; Black, R.E.; Campbell, H. Epidemiology and Etiology of Childhood Pneumonia in 2010: Estimates of Incidence, Severe Morbidity, Mortality, Underlying Risk Factors and Causative Pathogens for 192 Countries. *J. Glob. Health* **2013**, *3*, 010401. [PubMed]
3. Manikam, L.; Lakhanpaul, M. Epidemiology of Community Acquired Pneumonia. *Paediatr. Child Health* **2012**, *22*, 299–306. [CrossRef]
4. Rudan, I. Epidemiology and Etiology of Childhood Pneumonia. *Bull. World Health Organ.* **2008**, *86*, 408–416. [CrossRef]
5. Pneumonia in Children Statistics. Available online: <https://data.unicef.org/topic/child-health/pneumonia/> (accessed on 11 September 2024).
6. Jain, S.; Williams, D.J.; Arnold, S.R.; Ampofo, K.; Bramley, A.M.; Reed, C.; Stockmann, C.; Anderson, E.J.; Grijalva, C.G.; Self, W.H.; et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Children. *N. Engl. J. Med.* **2015**, *372*, 835–845. [CrossRef] [PubMed]
7. Gereige, R.S.; Laufer, P.M. Pneumonia. *Pediatr. Rev.* **2013**, *34*, 438–456. [CrossRef]
8. Eslamy, H.K.; Newman, B. Pneumonia in Normal and Immunocompromised Children: An Overview and Update. *Radiol. Clin.* **2011**, *49*, 895–920. [CrossRef] [PubMed]
9. Lee, K.H.; Gordon, A.; Foxman, B. The Role of Respiratory Viruses in the Etiology of Bacterial Pneumonia: An Ecological Perspective. *Evol. Med. Public Health* **2016**, *2016*, 95–109. [CrossRef] [PubMed]
10. Scotta, M.C.; Marostica, P.J.C.; Stein, R.T. 25—Pneumonia in Children. In *Kendig's Disorders of the Respiratory Tract in Children*, 9th ed.; Wilmott, R.W., Deterding, R., Li, A., Ratjen, F., Sly, P., Zar, H.J., Bush, A., Eds.; Elsevier: Philadelphia, PA, USA, 2019; pp. 427–438.e4. ISBN 978-0-323-44887-1.
11. Barson, W.J.; Kaplan, S.; Torchia, M. *Pneumonia in Children: Epidemiology, Pathogenesis, and Etiology*; UpToDate: Waltham, MA, USA, 2014.

12. Scott, J.A.G.; Wonodi, C.; Moïsi, J.C.; Deloria-Knoll, M.; DeLuca, A.N.; Karron, R.A.; Bhat, N.; Murdoch, D.R.; Crawley, J.; Levine, O.S. The Definition of Pneumonia, the Assessment of Severity, and Clinical Standardization in the Pneumonia Etiology Research for Child Health Study. *Clin. Infect. Dis.* **2012**, *54*, S109–S116. [[CrossRef](#)] [[PubMed](#)]
13. Kevat, P.M.; Morpeth, M.; Graham, H.; Gray, A.Z. A Systematic Review of the Clinical Features of Pneumonia in Children Aged 5–9 Years: Implications for Guidelines and Research. *J. Glob. Health* **2022**, *12*, 10002. [[CrossRef](#)] [[PubMed](#)]
14. Yao, D.; Xu, Z.; Lin, Y.; Zhan, Y. Accurate and Intelligent Diagnosis of Pediatric Pneumonia Using X-Ray Images and Blood Testing Data. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1058888. [[CrossRef](#)] [[PubMed](#)]
15. Scanlon, G.T.; Unger, J.D. The radiology of bacterial and viral pneumonias. *Radiol. Clin. North Am.* **1973**, *11*, 317–338. [[CrossRef](#)]
16. Grafakou, O.; Moustaki, M.; Tsoia, M.; Kavazarakis, E.; Mathioudakis, J.; Fretzayas, A.; Nicolaidou, P.; Karpathios, T. Can Chest X-Ray Predict Pneumonia Severity? *Pediatr. Pulmonol.* **2004**, *38*, 465–469. [[CrossRef](#)] [[PubMed](#)]
17. Kunz, W.G.; Patzig, M.; Crispin, A.; Stahl, R.; Reiser, M.F.; Notohamiprodjo, M. The Value of Supine Chest X-Ray in the Diagnosis of Pneumonia in the Basal Lung Zones. *Acad. Radiol.* **2018**, *25*, 1252–1256. [[CrossRef](#)]
18. Crame, E.; Shields, M.D.; McCrossan, P. Paediatric Pneumonia: A Guide to Diagnosis, Investigation and Treatment. *Paediatr. Child Health* **2021**, *31*, 250–257. [[CrossRef](#)]
19. Kazi, S.; Hernstadt, H.; Abo, Y.-N.; Graham, H.; Palmer, M.; Graham, S.M. The Utility of Chest X-Ray and Lung Ultrasound in the Management of Infants and Children Presenting with Severe Pneumonia in Low-and Middle-Income Countries: A Pragmatic Scoping Review. *J. Glob. Health* **2022**, *12*, 10013. [[CrossRef](#)]
20. Pereda, M.A.; Chavez, M.A.; Hooper-Miele, C.C.; Gilman, R.H.; Steinhoff, M.C.; Ellington, L.E.; Gross, M.; Price, C.; Tielsch, J.M.; Checkley, W. Lung Ultrasound for the Diagnosis of Pneumonia in Children: A Meta-Analysis. *Pediatrics* **2015**, *135*, 714–722. [[CrossRef](#)]
21. Wang, G.; Liu, X.; Shen, J.; Wang, C.; Li, Z.; Ye, L.; Wu, X.; Chen, T.; Wang, K.; Zhang, X.; et al. A Deep-Learning Pipeline for the Diagnosis and Discrimination of Viral, Non-Viral and COVID-19 Pneumonia from Chest X-Ray Images. *Nat. Biomed. Eng.* **2021**, *5*, 509–521. [[CrossRef](#)] [[PubMed](#)]
22. Khan, W.; Zaki, N.; Ali, L. Intelligent Pneumonia Identification From Chest X-Rays: A Systematic Literature Review. *IEEE Access* **2021**, *9*, 51747–51771. [[CrossRef](#)]
23. Udbhav, M.; Attri, R.K.; Vijarana, M.; Gupta, S.; Tripathi, K. Pneumonia Detection Using Chest X-Ray with the Help of Deep Learning. In *Concepts of Artificial Intelligence and its Application in Modern Healthcare Systems*; CRC Press: Boca Raton, FL, USA, 2023; ISBN 978-1-00-333308-1.
24. Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *J. Healthc. Eng.* **2019**, *2019*, 4180949. [[CrossRef](#)]
25. Li, Y.; Zhang, Z.; Dai, C.; Dong, Q.; Badrigilan, S. Accuracy of Deep Learning for Automated Detection of Pneumonia Using Chest X-Ray Images: A Systematic Review and Meta-Analysis. *Comput. Biol. Med.* **2020**, *123*, 103898. [[CrossRef](#)] [[PubMed](#)]
26. Necir, H.; Rebai, Y.; Nada, F. Novel Transfer Learning Approach for Early Detection of Childhood Pneumonia. *Res. Sq.* **2023**. [[CrossRef](#)]
27. Kahwachi, W.T.; Saed, K.A. A Novel Architecture of Convolutional Neural Network to Diagnose COVID-19 Disease. *Math. Stat. Eng. Appl.* **2022**, *71*, 4831–4856.
28. Walia, I.S.; Srivastava, M.; Kumar, D.; Rani, M.; Muthreja, P.; Mohadikar, G. Pneumonia Detection Using Depth-Wise Convolutional Neural Network (DW-CNN). *EAI Endorsed Trans. Pervasive Health Technol.* **2020**, *6*, e5. [[CrossRef](#)]
29. Reis, H.C.; Turk, V. COVID-DSNet: A Novel Deep Convolutional Neural Network for Detection of Coronavirus (SARS-CoV-2) Cases from CT and Chest X-Ray Images. *Artif. Intell. Med.* **2022**, *134*, 102427. [[CrossRef](#)]
30. Khan, E.; Rehman, M.Z.U.; Ahmed, F.; Alfouzan, F.A.; Alzahrani, N.M.; Ahmad, J. Chest X-Ray Classification for the Detection of COVID-19 Using Deep Learning Techniques. *Sensors* **2022**, *22*, 1211. [[CrossRef](#)] [[PubMed](#)]
31. Sriporn, K.; Tsai, C.-F.; Tsai, C.-E.; Wang, P. Analyzing Lung Disease Using Highly Effective Deep Learning Techniques. *Healthcare* **2020**, *8*, 107. [[CrossRef](#)] [[PubMed](#)]
32. Mohammed, F.A.; Tune, K.K.; Assefa, B.G.; Jett, M.; Muhie, S. Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 699–735. [[CrossRef](#)]
33. Fahim, M.A.I.; Naznin, F.; Moni, M.A.; Islam, M.Z. Improved Transfer Learning Architecture to Classify COVID-19 Affected Chest X-Rays Using Noisy Student Pre-Training. In Proceedings of the 2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Virtual, 16–19 August 2021; pp. 1–8.
34. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131.e9. [[CrossRef](#)]

35. Davila, M.H.; Jose Murillo, J.; Calisto, M.B.; Puente-Mejia, B.; Navarrete, D.; Riofrío, D.; Pérez, N.; Benítez, D.; Moyano, R.F. Measuring the Impact of Data Augmentation Techniques in Lung Radiograph Classification Using a Fractional Factorial Design: A COVID-19 Case Study. In Proceedings of the 2022 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI), Cali, Colombia, 27–29 July 2022; pp. 1–6.
36. Investigating Image Augmentation for Classification of Chest X-Ray Images | IEEE Conference Publication | IEEE Xplore. Available online: <https://ieeexplore.ieee.org/document/10008268> (accessed on 24 September 2024).
37. Effect of Augmented Datasets on Deep Convolutional Neural Networks Applied to Chest Radiographs—PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/31196565/> (accessed on 24 September 2024).
38. Team, K. Keras Documentation: Keras 3 API Documentation. Available online: <https://keras.io/api/> (accessed on 4 May 2024).
39. Module: Tf | TensorFlow v2.16.1. Available online: https://www.tensorflow.org/api_docs/python/tf (accessed on 4 May 2024).
40. Jagtap, A.D.; Kawaguchi, K.; Karniadakis, G.E. Adaptive Activation Functions Accelerate Convergence in Deep and Physics-Informed Neural Networks. *J. Comput. Phys.* **2020**, *404*, 109136. [\[CrossRef\]](#)
41. Qian, S.; Liu, H.; Liu, C.; Wu, S.; Wong, H.S. Adaptive Activation Functions in Convolutional Neural Networks. *Neurocomputing* **2018**, *272*, 204–212. [\[CrossRef\]](#)
42. Varshni, D.; Thakral, K.; Agarwal, L.; Nijhawan, R.; Mittal, A. Pneumonia Detection Using CNN Based Feature Extraction. In Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 20–22 February 2019; pp. 1–7.
43. Gupta, A.; Gupta, R.; Garg, N. An Efficient Approach for Classifying Chest X-Ray Images Using Different Embedder with Different Activation Functions in CNN. *J. Interdiscip. Math.* **2021**, *24*, 285–297. [\[CrossRef\]](#)
44. Sanida, T.; Sanida, M.V.; Sideris, A.; Dasygenis, M. Optimizing Lung Condition Categorization through a Deep Learning Approach to Chest X-Ray Image Analysis. *BioMedInformatics* **2024**, *4*, 2002–2021. [\[CrossRef\]](#)
45. Jang, J.; Cho, H.; Kim, J.; Lee, J.; Yang, S. Deep Neural Networks with a Set of Node-Wise Varying Activation Functions. *Neural Netw.* **2020**, *126*, 118–131. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Wang, S.-H.; Sakk, E. The Effect of Activation Function Choice on the Performance of Convolutional Neural Networks. *J. Emerg. Investig.* **2023**, *6*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for Activation Functions. *arXiv* **2017**, arXiv:1710.05941.
48. Misra, D. Mish: A Self Regularized Non-Monotonic Activation Function. *arXiv* **2019**, arXiv:1908.08681.
49. Sun, Y. The Role of Activation Function in Image Classification. In Proceedings of the 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), Beijing, China, 14–16 May 2021; pp. 275–278.
50. Alkhouly, A.A.; Mohammed, A.; Hefny, H.A. Improving the Performance of Deep Neural Networks Using Two Proposed Activation Functions. *IEEE Access* **2021**, *9*, 82249–82271. [\[CrossRef\]](#)
51. Nandi, A.; Jana, N.D.; Das, S. Improving the Performance of Neural Networks with an Ensemble of Activation Functions. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
52. Bingham, G.; Miiikkulainen, R. Discovering Parametric Activation Functions. *Neural Netw.* **2022**, *148*, 48–65. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Dunst, B.; Benedek, M.; Jauk, E.; Bergner, S.; Koschutnig, K.; Sommer, M.; Ischebeck, A.; Spinath, B.; Arendasy, M.; Bühner, M.; et al. Neural Efficiency as a Function of Task Demands. *Intelligence* **2014**, *42*, 22–30. [\[CrossRef\]](#)
54. Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *Neurocomputing* **2022**, *503*, 92–108. [\[CrossRef\]](#)
55. Nneji, G.U.; Cai, J.; Deng, J.; Monday, H.N.; James, E.C.; Ukwuoma, C.C. Multi-Channel Based Image Processing Scheme for Pneumonia Identification. *Diagnostics* **2022**, *12*, 325. [\[CrossRef\]](#)
56. Hammad, M.M. Deep Learning Activation Functions: Fixed-Shape, Parametric, Adaptive, Stochastic, Miscellaneous, Non-Standard, Ensemble. *arXiv* **2024**, arXiv:2407.11090.
57. Jaganathan, D.; Balsubramaniam, S.; Sureshkumar, V.; Dhanasekaran, S. Concatenated Modified LeNet Approach for Classifying Pneumonia Images. *J. Pers. Med.* **2024**, *14*, 328. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Gu, C.; Lee, M. Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-Ray Images. *Bioengineering* **2024**, *11*, 406. [\[CrossRef\]](#)
59. Manohar, B.; Das, R.; Lakshmi, M. A Hybridized LSTM-ANN-RSA Based Deep Learning Models for Prediction of COVID-19 Cases in Eastern European Countries. *Expert Syst. Appl.* **2024**, *256*, 124977. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Pediatric Pneumonia Recognition Using an Improved DenseNet201 Model with Multi-Scale Convolutions and Mish Activation Function

Petra Radočaj¹, Dorijan Radočaj^{2,*}  and Goran Martinović³ ¹ Layer d.o.o., Vukovarska cesta 31, 31000 Osijek, Croatia; radočajpetra@gmail.com² Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia³ Faculty of Electrical Engineering, Computer Science and Information Technology, Josip Juraj Strossmayer, University of Osijek, Kneza Trpimira 2B, 31000 Osijek, Croatia; goran.martinovic@ferit.hr

* Correspondence: dradočaj@fazos.hr; Tel.: +385-31-554-965

Abstract: Pediatric pneumonia remains a significant global health issue, particularly in low- and middle-income countries, where it contributes substantially to mortality in children under five. This study introduces a deep learning model for pediatric pneumonia diagnosis from chest X-rays that surpasses the performance of state-of-the-art methods reported in the recent literature. Using a DenseNet201 architecture with a Mish activation function and multi-scale convolutions, the model was trained on a dataset of 5856 chest X-ray images, achieving high performance: 0.9642 accuracy, 0.9580 precision, 0.9506 sensitivity, 0.9542 F1 score, and 0.9507 specificity. These results demonstrate a significant advancement in diagnostic precision and efficiency within this domain. By achieving the highest accuracy and F1 score compared to other recent work using the same dataset, our approach offers a tangible improvement for resource-constrained environments where access to specialists and sophisticated equipment is limited. While the need for high-quality datasets and adequate computational resources remains a general consideration for deep learning applications, our model's demonstrably superior performance establishes a new benchmark and offers the delivery of more timely and precise diagnoses, with the potential to significantly enhance patient outcomes.



Academic Editor: Maryam Ravan

Received: 2 January 2025

Revised: 2 February 2025

Accepted: 7 February 2025

Published: 10 February 2025

Citation: Radočaj, P.; Radočaj, D.; Martinović, G. Pediatric Pneumonia Recognition Using an Improved DenseNet201 Model with Multi-Scale Convolutions and Mish Activation Function. *Algorithms* **2025**, *18*, 98. <https://doi.org/10.3390/a18020098>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: convolutional neural network; deep learning; pediatric pneumonia; Mish activation function; multi-scale convolution

1. Introduction

Pediatric pneumonia constitutes a significant global health challenge, with a disproportionate impact on low- and middle-income countries (LMICs) [1], where it persists as a primary cause of mortality in children under five years of age [2]. According to the World Health Organization (WHO), pneumonia is responsible for approximately 14% of all fatalities in children under five years old; this translates to an estimated 740,000 deaths each year [3]. Globally, the disease accounts for 10–20 million hospitalizations and an annual incidence of 150–156 million cases, with over 80% of pneumonia-related deaths occurring in LMICs. Regions such as South Asia and West and Central Africa bear the highest burden, with pneumonia affecting approximately 1 in 71 children annually [4,5]. This infectious disease, characterized by inflammation of the pulmonary alveoli in one or both lungs, presents clinically in children with a spectrum of signs and symptoms, including cough, fever, dyspnea, and chest pain [6]. In severe presentations, the disease can progress to

hypoxemia (decreased arterial oxygen saturation) and potentially life-threatening complications such as pleural effusion and respiratory failure [7]. The etiology of pediatric pneumonia is diverse, encompassing bacterial, viral, and fungal pathogens. While the clinical presentation may exhibit variability based on the specific causative agent, the disease typically manifests with respiratory distress and systemic symptoms, necessitating timely diagnosis and appropriate therapeutic intervention [8,9]. In high-income countries, the mortality rates are significantly lower due to better access to vaccines, diagnostic tools, and antibiotics. However, pneumonia continues to impose a significant burden on healthcare systems worldwide, accounting for millions of outpatient visits and hospital admissions annually, as well as substantial healthcare costs. The financial burden of prolonged care is particularly severe in resource-constrained areas, where limited access to affordable healthcare exacerbates the issue [10,11]. Addressing this challenge requires not only improvements in prevention and treatment but also innovative diagnostic methods to ensure early and accurate detection [12].

Deep learning has become a pivotal tool in the diagnosis of pediatric pneumonia, offering advanced capabilities for analyzing large datasets of medical images. These algorithms can identify subtle patterns and anomalies that may elude human interpretation, providing consistent and objective diagnostic outcomes [13]. By reducing diagnostic errors and operating with greater efficiency than traditional methods, deep learning addresses key limitations of conventional approaches. Specifically, conventional methods often rely on manual interpretation, which can be prone to human error and variability, particularly when identifying subtle or atypical patterns in medical images. Deep learning overcomes these challenges by leveraging automated feature extraction and pattern recognition, enabling consistent and reproducible results. Additionally, traditional diagnostic methods can struggle with large volumes of data or complex datasets, whereas deep learning excels at processing and analyzing such data efficiently [14]. Its utility is particularly pronounced in resource-constrained settings, where access to skilled radiologists and advanced diagnostic facilities is limited, thereby promoting equitable healthcare delivery and enabling timely, accurate diagnoses in underserved populations [15,16]. CNNs, a specialized architecture within deep learning, have demonstrated remarkable accuracy in medical imaging. They excel at extracting and analyzing both local and global features from radiographic images, which are essential for identifying pathologies such as pneumonia [17]. In pediatric applications, deep learning models trained on extensive radiographic datasets can reliably detect features indicative of pneumonia, including those associated with early-stage or atypical presentations. This capability not only enhances diagnostic efficiency; however, it also serves as a valuable supplementary tool for clinicians—minimizing the risk of missed diagnoses and contributing to improved patient outcomes [18]. Although these advancements are significant, they still require ongoing refinement and validation in clinical settings because the stakes in medical diagnostics are exceptionally high.

Despite the considerable advancements in deep learning applications within the realm of medical imaging, existing models exhibit numerous limitations. Many of the leading models—such as Inception, ResNet, and VGG—are computationally demanding, necessitating extensive memory and processing capabilities [19,20]. These prerequisites render them difficult to implement on edge devices or in settings with limited resources [19]. Furthermore, these models frequently possess a vast number of parameters, which heightens the risk of overfitting when trained on small or imbalanced datasets; this scenario is particularly prevalent in medical imaging [21]. Another significant challenge pertains to the interpretability of deep learning models. Although these algorithms can achieve high accuracy, their opaqueness complicates clinicians' ability to trust and effectively integrate them into their practice [22,23]. Addressing these limitations is crucial for facilitating the

widespread adoption of deep learning techniques in pediatric pneumonia diagnostics. This study intends to tackle the deficiencies of current deep learning models by developing an enhanced DenseNet201-based model for pediatric pneumonia recognition.

The primary aim of this study is to optimize the architecture of a deep-learning model for pediatric pneumonia diagnosis. This is accomplished by leveraging a pretrained DenseNet201 model and introducing multi-scale convolutional layers. These layers facilitate the model in capturing features at varying resolutions, thus enhancing its ability to identify pneumonia across diverse radiographic presentations. The model was evaluated using a publicly available dataset, which ensures transparency and reproducibility of the results. Furthermore, the Mish activation function was incorporated to provide smoother gradient propagation and improved convergence compared to traditional activation functions, such as ReLU. An essential aspect of the model's design is its parameter efficiency, achieved by minimizing the number of parameters in comparison to other pre-trained models. This design choice makes the model particularly suitable for deployment in resource-constrained environments where computational resources may be limited. However, by utilizing transfer learning and a publicly accessible dataset, we demonstrated the model's robustness and diagnostic accuracy across a variety of imaging scenarios. Although challenges remain, this approach significantly contributes to the field of pediatric pneumonia diagnosis.

The primary contributions of this study are as follows:

- This study leveraged a pretrained DenseNet201 model to develop a novel architecture optimized for pediatric pneumonia diagnosis. By introducing multi-scale convolutional layers, the model effectively captures features at varying resolutions, enhancing its diagnostic accuracy across diverse radiographic presentations.
- The incorporation of the Mish activation function provides smoother gradient propagation and improved convergence compared to traditional activation functions, such as ReLU, further enhancing the model's stability and generalization capabilities.
- An essential focus was placed on parameter efficiency. The model achieves a significantly lower parameter count compared to other pretrained architectures, ensuring computational efficiency and making it ideal for deployment in resource-constrained settings where advanced hardware is limited.
- By employing transfer learning and a publicly accessible dataset, the study demonstrates the adaptability of the model for diverse imaging scenarios, addressing the diagnostic challenges faced in both high-resource and low-resource clinical environments.

This paper is structured as follows: Section 2 reviews the relevant literature on pediatric pneumonia diagnosis using deep learning techniques. Section 3 provides a detailed description of the proposed model and methodologies. Section 4 presents the experimental results, including a comparative analysis of model performance. Finally, Section 5 outlines the conclusions and directions for future research.

2. Related Works

Although deep learning has made significant progress in advancing automated pneumonia diagnosis, achieving higher performance and ensuring the scalability and robustness of these models in real-world clinical settings remain ongoing challenges. Previous studies have demonstrated moderate success in refining deep learning techniques for detecting pneumonia from medical imaging data.

Ha Pham and Tran [24] evaluated the performance of an ensemble model combining three CNN architectures (InceptionResNetV2, DenseNet201, and VGG16) for binary pneumonia classification using 5856 chest X-ray images. Compared to single CNN models, the ensemble demonstrated superior performance, achieving an accuracy of nearly 0.95 and improving the average F1 score by 3%. Additionally, the study underscores the potential

of ensemble methods to outperform individual architectures in medical image analysis tasks. Kahwachi and Saed [25] emphasized the importance of rapid COVID-19 diagnosis due to its widespread social and economic impact and the limitations of molecular tests. Their study utilized six CNN architectures (VGG19, InceptionV3, ResNet50, InceptionResNetV2, and DenseNet121) to classify chest CT images as COVID-19 or non-COVID-19, comparing the effects of various activation functions. They found that InceptionV3 and DenseNet121 performed best when paired with non-ReLU activation functions, highlighting their potential to improve diagnostic accuracy in resource-limited areas. Lujan-Garcia et al. [26] highlighted pneumonia as a leading cause of mortality in children under five and emphasized the importance of chest X-ray imaging for its diagnosis. In research, they used transfer learning with the ImageNet pre-trained Xception Network to classify chest X-rays, distinguishing between 3883 pneumonia cases and 1349 normal images. The model demonstrated competitive performance compared to state-of-the-art methods, achieving a precision of 0.84, a recall of 0.99, an F1 score of 0.91, and an AUC of 0.97. Jain et al. [27] highlighted pneumonia as a leading cause of death among children, with 880,000 deaths reported in 2016, particularly in South Asia and Sub-Saharan Africa. Using CNNs, the researchers classified chest X-ray images to classify pneumonia. They experimented with different parameters, hyperparameters, and the number of convolutional layers within the CNN architecture. Two custom CNN models achieved validation accuracies of 0.8526 and 0.9231, while pre-trained models like VGG16, VGG19, ResNet50, and InceptionV3 attained accuracies of 0.8728, 0.8846, 0.7756, and 0.7099, respectively. Kaya [28] emphasized the importance of rapid pediatric pneumonia detection due to its seasonal association and potentially fatal outcomes. The study utilized deep tuning transfer learning of state-of-the-art CNN models, specifically DenseNet121, achieving an accuracy of 0.9503 and an F1 score of 0.9603 on a dataset of 5856 chest X-ray images. The results demonstrated the effectiveness of the approach for accurate and timely pneumonia detection in pediatric cases. Wang et al. [29] proposed a DenseNet-based method with a Squeeze and Excitation (SE) block and max-pooling to enhance pneumonia classification by better focusing on lesion regions. They selected PReLU as the activation function to prevent neuron death during training. The model outperformed DenseNet, achieving 0.928 accuracy, 0.926 precision, 0.962 recall, and 0.943 F1 score, with improvements in all metrics.

Previous studies on pediatric pneumonia recognition using CNNs face limitations, including inadequate feature extraction for detecting subtle patterns in chest X-rays. For example, while ensemble models combining architectures like DenseNet201, InceptionResNetV2, and VGG16 demonstrated superior performance with improved F1 scores [24], their computational complexity makes them less practical for resource-constrained environments. Similarly, pre-trained CNNs such as VGG16, ResNet50, and InceptionV3 achieved validation accuracies between 0.7099 and 0.9231 in pediatric pneumonia recognition, but their reliance on extensive fine-tuning and higher resource requirements may limit their scalability in real-world applications [27]. Furthermore, despite promising results from DenseNet-based approaches incorporating SE blocks and custom activation functions [29], these methods often involve added architectural complexity, making deployment in clinical workflows challenging. Notably, many of these studies, including those by Kaya [28], Lujan-Garcia et al. [26], and Ha Pham and Tran [24], used the same publicly available Chest X-ray Images dataset as our study, highlighting dataset-specific limitations such as potential biases, variability in image quality, and limited demographic diversity. To tackle these deficiencies, this study introduces an innovative CNN model grounded in the DenseNet architecture. It integrates multi-scale convolution and the Mish activation function to improve feature detection, as well as gradient flow. The model achieves significant accuracy, sensitivity, and specificity; thus, it demonstrates a well-rounded performance in pneumonia

recognition, thereby avoiding misdiagnoses. This advancement effectively reduces false negatives, which can lead to overlooked diagnoses, and minimizes false positives, thus preventing unnecessary treatments. Although the model enhances scalability, robustness, and clinical applicability, it ultimately serves as a reliable tool to aid healthcare professionals in diagnosing pediatric pneumonia, particularly in low-resource settings.

3. Materials and Methods

We introduce a novel approach that leverages the DenseNet architecture, augmented with the Mish activation function and multi-scale convolutions (see Supplementary Materials). Through transfer learning, this method aims to enhance the recognition of pediatric pneumonia. The proposed methodology comprises three primary stages, as depicted in Figure 1: (1) preprocessing and preparation of input data, including the categorization of chest X-ray images into healthy and pneumonia classes; (2) implementation of the designed deep learning model, incorporating the DenseNet architecture with the Mish activation function and multi-scale convolutions; and (3) quantitative evaluation of the model's diagnostic performance in pneumonia detection, emphasizing metrics such as accuracy, precision, F1 score, sensitivity, and specificity.

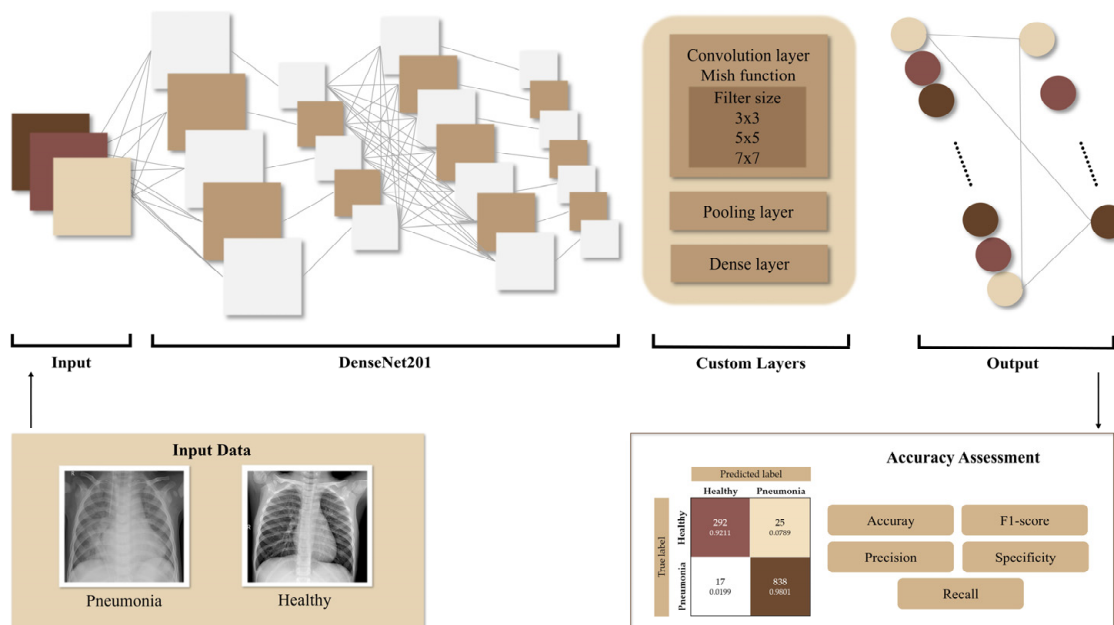


Figure 1. The study workflow for pneumonia recognition, utilizing a DenseNet architecture, Mish activation function, and multi-scale convolutions, proceeded through the following phases: (1) input data preparation, encompassing both healthy and pneumonia image classes; (2) implementation of the proposed deep learning model; and (3) evaluation of the implemented model's performance in pneumonia recognition, including accuracy assessment.

3.1. Data Preparation and Experimental Setup

We used the publicly available Chest X-ray Images (Pneumonia) dataset [30] for our experiments, which includes X-ray images of pediatric patients aged one to five years from Guangzhou Women and Children's Medical Center. The dataset contains 5856 images, categorized into two groups: pneumonia (4273 images) and healthy (1583 images) [30]. All chest radiographs were initially screened for quality control, and low-quality or unreadable scans were removed prior to analysis. Diagnoses for the images were graded by two expert physicians, with a third expert verifying the evaluation set to minimize grading errors and ensure the reliability of the dataset [30]. Samples of used images are shown in Figure 2. To ensure a balanced evaluation, we performed a stratified random split of the dataset into

training and validation subsets in an 80:20 ratio, preserving the class distribution. Due to computational resource constraints and the large dataset size, we opted for a single 80:20 split instead of k-fold cross-validation, which would have significantly increased training time without providing proportional benefits in this context. Although we did not apply explicit measures such as class weighting or oversampling to address the class imbalance, the CNN inherently adapts to the data distribution through its training process. Additionally, data augmentation techniques, including rotation, shifting, zooming, and horizontal flipping, were applied during preprocessing to increase dataset variability and improve model generalization. All images were resized to 224×224 pixels.

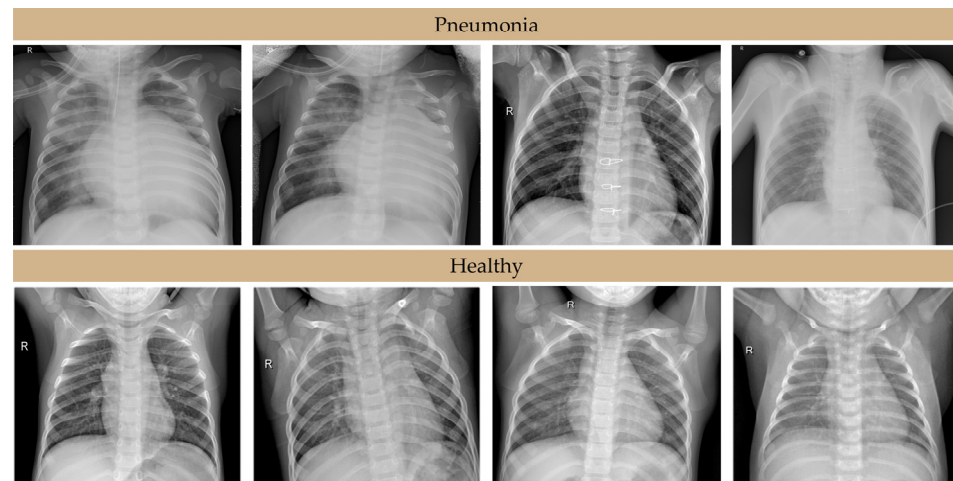


Figure 2. Samples of the pneumonia and healthy chest X-ray images used in this study.

We developed and trained the deep learning models using Python with the Keras-GPU [31] and TensorFlow-GPU [32] frameworks on the Google Colab platform. The platform's NVIDIA Tesla K80 GPUs with 12 GB of memory enabled efficient training. The models were trained over 20 epochs with a batch size of 32, utilizing the Adam optimization algorithm to update model parameters dynamically. The learning rate was automatically adjusted during training using a reduction mechanism that lowered it when validation performance plateaued, with a minimum of 0.5×10^{-6} to optimize convergence and prevent overfitting.

3.2. Model Architecture for Proposed Approach

Multi-scale convolution is a powerful technique in CNNs that enhances feature extraction by capturing information across varying spatial resolutions. Unlike standard convolutional layers that operate with fixed filter sizes, multi-scale convolution utilizes filters of different dimensions to simultaneously extract fine-grained and coarse-grained features [33]. This capability is particularly beneficial for medical image analysis, such as pediatric pneumonia recognition, where diverse patterns and textures must be identified across varying spatial scales. By combining outputs from filters of varying sizes, our model effectively captures both local and global features, making it more robust to variations in image resolution and scale [34,35]. This ability to capture multi-resolution features improves the model's generalizability, particularly in clinical datasets with heterogeneous imaging conditions [36,37]. Additionally, multi-scale convolution reduces the risk of overfitting by diversifying the spatial patterns used during feature extraction.

In this research, we utilized the pretrained DenseNet201 network as the foundational framework for feature extraction. DenseNet201, renowned for its intricately connected structure and effective feature reuse, was pretrained on ImageNet, thus providing a substantial basis for transfer learning. We chose to freeze the pretrained layers of DenseNet201

to preserve the learned features; however, we incorporated multi-scale convolutional layers on top. Specifically, we implemented filters of sizes 3×3 , 5×5 , and 7×7 , which enabled the model to capture features across various spatial resolutions. These multi-scale outputs were then concatenated, forming a more comprehensive feature representation, which improved the model's capacity to detect pneumonia-related patterns in chest X-ray images. Additionally, we enhanced the architecture with pooling operations and a fully connected dense layer for classification because this allows for more accurate predictions. The convolutional operation in this architecture is expressed in Equation (1):

$$y(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(i+m, j+n) * w(m, n), \quad (1)$$

where $x(i+m, j+n)$ represents the input feature map, $w(m, n)$ denotes the convolution filter of size $M \times N$, and $y(i, j)$ is the resulting output feature [38]. The multi-scale convolutional framework incorporates filters of different dimensions to capture diverse spatial patterns, thereby enhancing the model's ability to recognize complex and heterogeneous radiographic presentations, as demonstrated in Algorithm 1.

Algorithm 1: Multi-Scale Convolution

```

1.  function MultiScaleConvolution (Input_Image, Kernels, Biases)
2.      Input:
3.          Input_Image: 2D matrix of pixel intensities
4.          Kernels: List of 2D filter matrices [Kernel1, Kernel2, ..., KernelN]
5.          Biases: List of corresponding bias terms [Bias1, Bias2, ..., BiasN]
6.      Output: Combined_Feature_Map
7.      InitializeFeature_Maps as an empty list
8.      for each scale  $k$  in range 1,  $N$  do
9.          Kernel  $\leftarrow$  Kernels[ $k$ ]
10.         Bias  $\leftarrow$  Biases[ $k$ ]
11.         Feature_Map $k$   $\leftarrow$  Convolution (Input_Image, Kernel, Bias)
12.         Append Feature_Map $k$  to Feature_Maps
13.      end for
14.      Combine all Feature_Map $k$  into a single representation:
15.      Combined_Feature_Map  $\leftarrow$  Concatenate (Feature_Maps)
16.      return Combined_Feature_Map
17.  end function

```

To further improve the model's efficacy, the Mish activation function was utilized in the multi-scale convolutional layers. Mish, as mathematically defined in Equation (2):

$$\text{Mish}(x) = x * \tanh(\text{softplus}(x)) = x * \tanh(\ln(1 + e^x)) \quad (2)$$

provides smoother gradient propagation and better convergence compared to traditional activation functions such as ReLU [39]. Unlike ReLU, which truncates negative values, Mish retains negative information through a smooth, non-monotonic curve, improving the model's capacity to learn subtle patterns and enhancing its generalization capabilities [39,40]. Supporting its adoption in this work, prior studies show Mish's superior performance compared to ReLU and Swish in CNN-based tasks, particularly within the field of medical imaging, including applications for pediatric pneumonia recognition [41]. Moreover, we conducted experiments comparing our model using the Mish activation function against the same model using ReLU. This comparative analysis allowed us to em-

pirically evaluate the impact of Mish on performance metrics and validate its effectiveness in our specific application.

Overall, the proposed approach is demonstrated in Algorithm 2.

Algorithm 2: Multi-Scale Convolution and Mish activation function with Pretrained DenseNet201

1. **Step 1: Define Mish Activation Function**
 2. Define the Mish activation function:
 3. $Mish(x) = x * \tanh(\text{softplus}(x))$, where:
 4. $\text{softplus}(x) = \log(1 + \exp(x))$,
 5. $\tanh(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x))$.
 6. **Step 2: Load Pretrained Model**
 7. Load DenseNet201 pretrained on ImageNet without the top layers.
 8. Freeze the pretrained layers for feature extraction.
 9. **Step 3: Add Multi-Scale Convolutions**
 10. Extract features from DenseNet201 (*base_model.output*).
 11. Apply multi-scale convolutional layers:
 12. *conv_small* with filter size 2×2 , activation: *mish*.
 13. *conv_medium* with filter size 4×4 , activation: *mish*.
 14. *conv_large* with filter size 8×8 , activation: *mish*.
 15. Concatenate multi-scale features.
 16. **Step 4: Global Pooling and Classification**
 17. Apply *GlobalMaxPooling2D* on concatenated features.
 18. Add a Dense layer with sigmoid activation for binary classification.
 19. **Step 5: Compile and Train the Model**
 20. Compile the model using:
 21. Optimizer: *Adam*.
 22. Loss function: *binary_crossentropy*.
 23. Metrics: *Accuracy and AUC*.
 24. Train the model with the training generator.
-

3.3. Pediatric Pneumonia Performance Evaluation

We assessed the performance of deep learning models for classifying pneumonia in chest X-ray images using precision, recall, specificity, F1 score, and accuracy, as defined in Equations (3)–(7). These metrics, calculated from the confusion matrix, provided a comprehensive evaluation of the models' diagnostic capabilities:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (5)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (6)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

We defined true positives (TP) as cases where the model correctly identified pneumonia and true negatives (TN) as cases without pneumonia that were accurately classified. False

positives (FP) occurred when the model incorrectly classified non-pneumonia cases as pneumonia, and false negatives (FN) represented missed pneumonia cases. Specificity measured the model's ability to correctly identify non-pneumonia cases, complementing recall, which focused on detecting pneumonia cases. Chest X-ray images with pneumonia typically showed patterns such as increased lung density, consolidation, or pleural effusion, while diagnostic errors often resulted from overlapping features with other conditions or artifacts in the images.

4. Results and Discussion

The proposed model with the Mish activation function achieved notable performance in the diagnosis of pediatric pneumonia from radiological images, demonstrating an overall accuracy of 0.9642. A more comprehensive evaluation of accuracy assessment is presented in Table 1. This high accuracy indicates the model's effectiveness in correctly classifying both pneumonia-positive and healthy samples within the dataset, as reflected in the associated performance metrics. Furthermore, the table presented the results of the ablation study, comparing the Mish activation function against ReLU. This comparison provided the numerical evidence mentioned earlier, quantifying the impact of Mish on performance metrics and validating its effectiveness in this specific application. The F1 score was employed as a key evaluation metric due to its ability to provide a harmonic mean of precision and recall, particularly valuable in datasets characterized by class imbalance. During the research evaluation, it was concluded that accuracy is the most commonly used metric. We utilized it to enable an objective comparison with other studies. In addition to accuracy, the F1 score was also employed, although it was observed that many studies did not include this metric despite dealing with imbalanced datasets. To encourage future studies to adopt a fully objective approach to evaluating solutions, we emphasized the following metrics in our analysis: Accuracy, Precision, Recall (sensitivity), F1 score, and specificity, thereby ensuring a comprehensive and detailed evaluation. In this study, we employed a dataset consisting of 4273 chest X-ray images of pneumonia cases and 1583 images of healthy individuals. Despite the observed class imbalance, the model achieved a high F1 score of 0.9542, indicating a negligible impact of imbalance on the model's capacity for accurate classification. We extensively evaluated sensitivity and specificity to ensure accurate diagnoses, minimizing both missed cases and unnecessary treatments and ultimately improving patient outcomes. The model demonstrated a sensitivity of 0.9506, reflecting its ability to correctly identify pediatric patients with pneumonia while reducing the risk of missed or delayed diagnoses, which can have a serious impact on patient outcomes. Onwards, the model also demonstrated a high specificity of 0.9507, reflecting its ability to accurately classify healthy patients and to minimize false positives, thereby reducing unnecessary treatments such as inappropriate antibiotic use, which can lead to unwanted side effects and contribute to antibiotic resistance.

Table 1. Accuracy assessment of the proposed approach.

Metric	Model with Mish	Model with ReLU
Accuracy	0.9642	0.9616
Precision	0.9580	0.9534
Recall (Sensitivity)	0.9506	0.9489
F1 score	0.9542	0.9511
Specificity	0.9507	0.9489

These findings are corroborated by the confusion matrix, which is shown in Figure 3. The model correctly identified 838 pneumonia-positive cases (true positives) and 292 healthy cases (true negatives) while misclassifying 25 healthy cases (false positives) and 17 pneumonia cases (false negatives). These outcomes align with the reported performance metrics, underscoring the model's reliability and effectiveness in pediatric pneumonia diagnosis. Further analysis of these misclassified instances suggested that some errors arose from cases exhibiting borderline radiological features or overlapping radiological characteristics with other pulmonary conditions, such as bronchitis or atelectasis. The findings suggest that augmenting the training dataset and implementing region-specific feature enhancement techniques could lead to a reduction in such misclassifications.

		Predicted label	
		Healthy	Pneumonia
True label	Healthy	292 0.9211	25 0.0789
	Pneumonia	17 0.0199	838 0.9801

Figure 3. The confusion matrix for the proposed methodology. The classified values are represented by the upper numerical entry in each cell, with the corresponding class-specific percentage shown below.

To further validate the model, enhance interpretability, and facilitate clinical integration, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed. Grad-CAM generates visual explanations by producing heatmaps that highlight the image regions most salient to the model's classification. In our architecture, this technique was applied to the last convolutional layer to ensure that the extracted features were both spatially and semantically rich. This technique is particularly valuable in the context of pediatric pneumonia diagnosis, as it offers transparency into the model's decision-making process, enabling verification of its diagnostic rationale and fostering clinical trust. In healthy cases, Grad-CAM typically generates diffuse activation patterns across the lung fields, consistent with the absence of focal pathology. Conversely, in cases of viral pneumonia, Grad-CAM heatmaps demonstrate more distributed and heterogeneous activation, often encompassing perihilar regions and reflecting the characteristic interstitial infiltrates associated with viral infection. In contrast, bacterial pneumonia exhibits more localized and intense activation patterns, concentrated over regions of consolidation, aligning with the typical lobar presentation of bacterial infection. The results are presented in Figure 4.

The proposed approach yielded an accuracy of 0.9642, placing it among the highest reported accuracies in this domain. A comparative evaluation is provided in Table 2. This result is slightly higher than the accuracy achieved by Vrbancic and Podgorelec [42], who reported an accuracy of 0.9626 using their SGDRE method on a similarly sized dataset. Other studies, such as Jaganathan et al. [43] using the LeNet-5 model on a significantly larger dataset of 84,484 images, reported an accuracy of 0.9600, which is comparable to ours despite the larger dataset. Notably, many methods using ensemble architectures, such as

Ha Pham and Tran [24], who reported an accuracy of 0.9503, and Mabrouk et al. [44], who reported an accuracy of 0.9391, achieved slightly lower accuracy values despite leveraging advanced combinations of models. This suggests that our single-model approach provides state-of-the-art performance without requiring the complexity of ensemble methods. In terms of the F1 score, our approach achieved 0.9640, outperforming most other methods that explicitly reported this metric. For example, Wang et al. [29], employing a custom CNN with a dataset of 5857 images, achieved an F1 score of 0.9430, while Mabrouk et al. [44] reported an F1 score of 0.9343 with their ensemble method on a dataset of the same size. Notably, Jaganathan et al. [43] and AlGhamdi [45] also achieved F1 scores of 0.9600 using LeNet-5 and MobileNetV3, respectively, with larger datasets. These comparisons highlight the efficiency of our approach, which performs competitively with methods that use significantly more training data. Some studies, such as Mardianto et al. [46], did not report F1 scores, focusing only on accuracy. This omission is concerning, as accuracy alone does not provide insight into the balance between false positives and false negatives, which is particularly important in medical diagnostics where class imbalances are common. By not reporting F1 scores, these studies risk underestimating the importance of false negative rates, which can have critical clinical implications. Reporting the F1 score, on the other hand, ensures a more comprehensive evaluation of a model's performance in such sensitive applications.

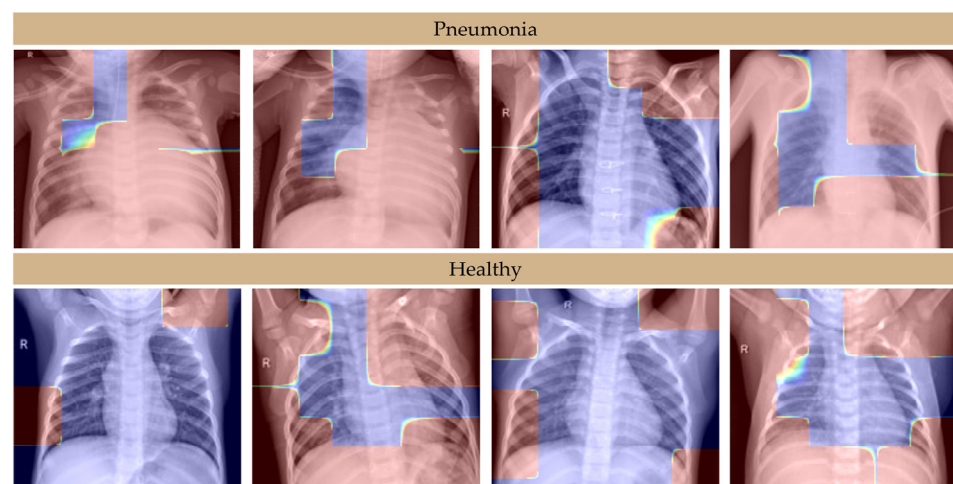


Figure 4. Grad-CAM heatmaps demonstrating model focus in pediatric pneumonia diagnosis.

Our approach, based on the improved DenseNet201 model with multi-scale convolutions and the Mish activation function, achieved superior metrics compared to both custom CNNs and hybrid combinations of CNNs with other methods. For instance, our accuracy of 0.9642 surpasses the performance of Wang et al. [29], who used a custom CNN which performed an accuracy of 0.9280 and F1 score of 0.9430, and Stephen et al. [22], who also employed a custom CNN but did not report an F1 score. Similarly, our results outperform Mardianto et al. [46], who combined a CNN with SVM and achieved an accuracy of 0.9200 without reporting an F1 score. Compared to ensemble methods, such as Ha Pham and Tran [24], our approach demonstrated a higher accuracy of 0.9642 compared to 0.9503 and similar F1 scores of 0.9640 compared to 0.9604. This suggests that our model is more streamlined while maintaining comparable or better performance than more complex architectures. Instead of developing custom CNNs, which demand extensive hyperparameter tuning and large training datasets for effective generalization, our approach leverages transfer learning with a pre-trained DenseNet201 model. The utilization of features pre-trained on the extensive ImageNet dataset facilitates efficient training and robust performance, even

with limited medical data, representing a substantial advantage in medical imaging where the acquisition of large, annotated datasets is often both challenging and costly. While custom CNNs trained on limited data are susceptible to overfitting and poor generalization, transfer learning with DenseNet201 provides a strong inductive bias, promoting better generalization and faster convergence. The incorporation of multi-scale convolutions and the Mish activation function further refine feature extraction from chest X-rays, surpassing standard transfer learning performance. This combined strategy of transfer learning with targeted architectural modifications offers a more efficient and effective alternative to building custom CNNs from scratch.

A clear progression in performance metrics can be observed over recent years as models have evolved to leverage larger datasets and more advanced architectures. For instance, early approaches, such as Manickam et al. [47] using ResNet50 in 2021, achieved an accuracy of 0.9306 and an F1 score of 0.9271 on 5232 images. Similarly, Mabrouk et al. [44] in 2022 achieved an accuracy of 0.9391 and an F1 score of 0.9343 using ensemble methods. However, more recent studies, such as Vrbancic and Podgorelec [42] in 2022, who reported an accuracy of 0.9626 accuracy and an F1 score of 0.974, and Kaya [28] in 2024, who reported an accuracy of 0.9503 accuracy and an F1 score of 0.9603, reflect a steady improvement in results, primarily due to advancements in model architectures and training strategies.

Table 2. Comparison and evaluation of the proposed approach and scientific studies in the last six years according to the data from the Web of Science Core Collection.

Reference	Year	CNN Model	No. of Images	Accuracy	F1 Score
Mardianto et al. [46]	2024	CNN+SVM	6140	0.9200	-
AlGhamdi [45]	2024	MobileNetV3	14,000>	-	0.9600
Jaganathan et al. [43]	2024	LeNet-5	84,484	0.9600	0.9600
Dzhaynakbaev et al. [48]	2024	VGG16	5228	-	-
Ha Pham and Tran [24]	2024	Ensemble (InceptionResNetV2, DenseNet201, VGG16)	5856	0.9503	0.9604
Mabrouk et al. [44]	2022	Ensemble (MobileNetV2, DenseNet169, Vision Transformer)	5856	0.9391	0.9343
Stephen et al. [22]	2019	Custom CNN	5856	0.9531	-
Lan et al. [49]	2024	DenseNet121	578	0.8100	-
Kaya [28]	2024	DenseNet121	5856	0.9503	0.9603
Manickam et al. [47]	2021	ResNet50	5232	0.9306	0.9271
Wang et al. [29]	2022	Custom CNN	5857	0.9280	0.9430
Vrbancic and Podgorelec [42]	2022	SGDRE method	5858	0.9626	0.9744
Yi et al. [50]	2023	Custom CNN	5856	0.9609	-
Our approach			5856	0.9642	0.9640

Hybrid models, such as the one proposed by Mardianto et al. [46], which integrate CNN with SVM, leverage the improved classification boundaries provided by SVM. However, their relatively lower accuracy of 0.9200 and omission of F1 scores raise concerns about their ability to manage class imbalances effectively. Similarly, ensemble methods such as those employed by Ha Pham and Tran [24] and Mabrouk et al. [44] leverage the diversity of models like InceptionResNetV2, DenseNet201, and Vision Transformers to

achieve competitive F1 scores of 0.9604 and 0.9343, respectively. Despite these advantages, the complexity and resource demands of ensemble architectures pose significant challenges in terms of interpretability, computational efficiency, and scalability. Custom CNNs, including those developed by Stephen et al. [22], Wang et al. [29], and Yi et al. [50], exhibit flexibility and adaptability to specific datasets, yielding accuracy values ranging from 0.9280 to 0.9609. Their performance highlights the potential of tailored architectures; however, the omission of F1 scores in some cases and slightly lower performance compared to pre-trained models suggest limited generalizability and challenges in capturing complex data patterns. Pre-trained architectures like DenseNet and ResNet demonstrate consistent performance due to their robust design. For example, Kaya [28] and Manickam et al. [47] reported accuracy values of 0.9503 and 0.9306, respectively, showcasing the efficacy of DenseNet121 and ResNet50 in feature propagation and handling gradient-related issues. However, DenseNet's performance can vary depending on dataset size and complexity, as evidenced by the low accuracy of 0.8100 achieved by Lan et al. [49] on a small dataset of 578 images. Among lightweight models, AlGhamdi [45] utilized MobileNetV3 to achieve an F1 score of 0.9600 on a larger dataset (>14,000 images), demonstrating its efficiency for resource-constrained environments. However, the absence of accuracy metrics limits a holistic evaluation of its performance. Similarly, foundational architectures like LeNet-5, used by Jaganathan et al. [43], achieved competitive accuracy and F1 scores of 0.9600 on a significantly larger dataset with 84484 images, though their simplicity may constrain their ability to capture complex features compared to modern architectures. Optimization-based approaches such as the SGDRE method by Vrbancic and Podgorelec [42] delivered an outstanding performance, achieving the highest F1 score of 0.9744 and an accuracy of 0.9626. While this underscores the potential of targeted optimization strategies, such methods might face limitations in adaptability across diverse datasets. The proposed approach, leveraging an improved DenseNet201 model enhanced with multi-scale convolutions and the Mish activation function, outperformed most compared methods, achieving the highest accuracy of 0.9642 and an F1 score of 0.9640. Its ability to deliver superior performance without requiring ensemble complexities highlights its efficiency. However, like other pre-trained architectures, its reliance on transfer learning may limit domain-specific interpretability compared to custom CNNs tailored exclusively for the task.

By utilizing a deep convolutional architecture, the model significantly reduces computational demands compared to ensemble-based methodologies, making it suitable for deployment in resource-constrained environments. This efficiency enables integration into point-of-care diagnostic tools, such as portable X-ray devices and mobile applications, thereby broadening access to accurate and timely diagnoses in low-resource and remote clinical settings. Additionally, the model has the potential to integrate seamlessly into telemedicine workflows, where portable X-ray systems combined with cloud-based analysis could provide rapid diagnostic support for patients in underserved areas. The model's resilience to class imbalance, in conjunction with training on publicly available datasets, significantly enhances its generalizability across varied patient populations. This model is also engineered to accommodate cloud-based deployment strategies, which yield several notable advantages. For instance, it offers scalability to manage substantial volumes of radiological data originating from multiple healthcare facilities. However, it also enhances accessibility, allowing clinicians in under-resourced areas to upload and analyze images in real time. Despite these strengths, limitations include the reliance on publicly available datasets, which may not fully capture the variability in image quality or geographic and demographic diversity. Plans for external validation using diverse real-world datasets are essential to further establish the model's reliability and applicability. Although it establishes a framework for continuous model refinement through iterative updates and retraining,

this ensures alignment with evolving diagnostic standards and expanding datasets. A possible solution that incorporates cloud-based infrastructure could facilitate centralized data processing, all while preserving the model's clinical relevance over time, making it an adaptable and sustainable solution for medical image analysis.

5. Conclusions and Future Work

In summary, the advanced deep learning model, which employs the DenseNet201 architecture, along with the Mish activation function and multi-scale convolutions, demonstrates impressive performance in diagnosing pediatric pneumonia from chest X-rays. Trained on a dataset comprising 5856 images, it achieved an overall accuracy of 96.42%, thereby securing its position among the top methodologies within this field. However, its performance not only exceeds that of numerous ensemble-based techniques, but it also retains a streamlined and computationally efficient design. This architectural efficiency is particularly advantageous for deployment in resource-constrained environments because ensemble methods can often be impractical due to their inherent complexity and computational requirements. The model's ability to provide timely and accurate diagnoses offers a scalable solution for addressing diagnostic gaps in low- and middle-income countries, where pneumonia remains a leading cause of childhood mortality, contributing to over 740,000 deaths annually. By reducing reliance on radiologists and advanced diagnostic infrastructure, the model has the potential to enhance healthcare delivery in under-resourced regions.

Despite its strengths, limitations include dependency on high-quality datasets, potential biases in training data, and challenges in generalizing across diverse populations. Furthermore, epidemiological and demographic distributions in LMICs may differ significantly from the public datasets used in this study. This limitation highlights the need for further validation using real-world and diverse demographic data to ensure broader applicability and reliability of the model. Future work could also explore multi-class tasks, such as distinguishing between bacterial and viral pneumonia, to enhance diagnostic precision and clinical relevance. To address these limitations, we aim to incorporate additional imaging modalities and integrate the model into clinical decision support systems. Scalability for other modalities, such as ultrasound, represents an important opportunity for expanding the model's utility across different healthcare contexts. Additionally, cloud-based deployment strategies could enhance scalability, enabling real-time analysis and continuous model refinement to adapt to evolving clinical needs. This approach represents a significant step forward in leveraging artificial intelligence for equitable and effective healthcare delivery.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/a18020098/s1>, Table S1. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update [51].

Author Contributions: Conceptualization, P.R.; methodology, P.R.; software, P.R.; validation, D.R. and G.M.; formal analysis, P.R.; investigation, P.R.; resources, P.R.; data curation, P.R.; writing—original draft preparation, P.R.; writing—review and editing, P.R., D.R. and G.M.; visualization, P.R.; supervision, G.M.; project administration, G.M.; funding acquisition, D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The code developed in this study is available upon request from the corresponding author. The open access Chest X-ray Images (Pneumonia) repository containing images collected from pediatric patients, is divided into two categories (Pneumonia/Healthy) and is available

at <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia> (accessed on 12 December 2024).

Conflicts of Interest: Author Petra Radočaj was employed by the company “Layer d.o.o.”. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Marangu, D.; Zar, H.J. Childhood Pneumonia in Low-and-Middle-Income Countries: An Update. *Paediatr. Respir. Rev.* **2019**, *32*, 3–9. [CrossRef] [PubMed]
2. Selvi, M.; Vaithilingan, S. Childhood Pneumonia in Low- and Middle-Income Countries: A Systematic Review of Prevalence, Risk Factors, and Healthcare-Seeking Behaviors. *Cureus* **2024**, *16*, e57636. [CrossRef]
3. Pneumonia in Children. Available online: <https://www.who.int/news-room/fact-sheets/detail/pneumonia> (accessed on 22 December 2024).
4. Rudan, I. Epidemiology and Etiology of Childhood Pneumonia. *Bull. World Health Organ.* **2008**, *86*, 408–416. [CrossRef] [PubMed]
5. Genie, Y.D.; Sayih, A.; Dessalegn, N.; Adugnaw, E.; Hiwot, A.Y.; Tesfa, T.B.; Kindie, K.; Gutema, L.; Ayalew, E.; Kebede, B.F. Time to Recovery from Severe Pneumonia and Its Predictors among Pediatric Patients Admitted in South West Region Governmental Hospitals, South West Ethiopia: Prospective Follow-up Study. *Glob. Pediatr.* **2024**, *9*, 100227. [CrossRef]
6. Kevat, P.M.; Morpeth, M.; Graham, H.; Gray, A.Z. A Systematic Review of the Clinical Features of Pneumonia in Children Aged 5–9 Years: Implications for Guidelines and Research. *J. Glob. Health* **2022**, *12*, 10002. [CrossRef] [PubMed]
7. Salah, E.T.; Algasim, S.H.; Mhamoud, A.S.; Husian, N.E.O.S.A. Prevalence of Hypoxemia in Under-Five Children with Pneumonia in an Emergency Pediatrics Hospital in Sudan. *Indian J. Crit. Care Med. Peer-Rev. Off. Publ. Indian Soc. Crit. Care Med.* **2015**, *19*, 203–207. [CrossRef] [PubMed]
8. Grief, S.N.; Loza, J.K. Guidelines for the Evaluation and Treatment of Pneumonia. *Prim. Care Clin. Off. Pract.* **2018**, *45*, 485–503. [CrossRef] [PubMed]
9. Zhang, J.; Zhu, Y.; Zhou, Y.; Gao, F.; Qiu, X.; Li, J.; Yuan, H.; Jin, W.; Lin, W. Pediatric Adenovirus Pneumonia: Clinical Practice and Current Treatment. *Front. Med.* **2023**, *10*, 1207568. [CrossRef] [PubMed]
10. Haggie, S.; Selvadurai, H.; Gunasekera, H.; Fitzgerald, D.A. Paediatric Pneumonia in High-Income Countries: Defining and Recognising Cases at Increased Risk of Severe Disease. *Paediatr. Respir. Rev.* **2021**, *39*, 71–81. [CrossRef]
11. Madhi, S.A.; De Wals, P.; Grijalva, C.G.; Grimwood, K.; Grossman, R.; Ishiwada, N.; Lee, P.-I.; Nascimento-Carvalho, C.; Nohynek, H.; O’Brien, K.L.; et al. The Burden of Childhood Pneumonia in the Developed World: A Review of the Literature. *Pediatr. Infect. Dis. J.* **2013**, *32*, e119–e127. [CrossRef]
12. Kanwal, K.; Asif, M.; Khalid, S.G.; Liu, H.; Qurashi, A.G.; Abdullah, S. Current Diagnostic Techniques for Pneumonia: A Scoping Review. *Sensors* **2024**, *24*, 4291. [CrossRef] [PubMed]
13. Li, Y.; Zhang, Z.; Dai, C.; Dong, Q.; Badrigilan, S. Accuracy of Deep Learning for Automated Detection of Pneumonia Using Chest X-Ray Images: A Systematic Review and Meta-Analysis. *Comput. Biol. Med.* **2020**, *123*, 103898. [CrossRef] [PubMed]
14. Hameed, B.M.Z.; Prerepa, G.; Patil, V.; Shekhar, P.; Zahid Raza, S.; Karimi, H.; Paul, R.; Naik, N.; Modi, S.; Vigneswaran, G.; et al. Engineering and Clinical Use of Artificial Intelligence (AI) with Machine Learning and Data Science Advancements: Radiology Leading the Way for Future. *Ther. Adv. Urol.* **2021**, *13*, 17562872211044880. [CrossRef] [PubMed]
15. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdass, M.; Kern, C.; et al. A Comparison of Deep Learning Performance against Health-Care Professionals in Detecting Diseases from Medical Imaging: A Systematic Review and Meta-Analysis. *Lancet Digit. Health* **2019**, *1*, e271–e297. [CrossRef]
16. Guo, Y.; Huang, C.; Sheng, Y.; Zhang, W.; Ye, X.; Lian, H.; Xu, J.; Chen, Y. Improve the Efficiency and Accuracy of Ophthalmologists’ Clinical Decision-Making Based on AI Technology. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 192. [CrossRef] [PubMed]
17. Varshni, D.; Thakral, K.; Agarwal, L.; Nijhawan, R.; Mittal, A. Pneumonia Detection Using CNN Based Feature Extraction. In Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 20–22 February 2019; pp. 1–7.
18. Rajaraman, S.; Thoma, G.; Antani, S.; Candemir, S. Visualizing and Explaining Deep Learning Predictions for Pneumonia Detection in Pediatric Chest Radiographs. In *Medical Imaging 2019: Computer-Aided Diagnosis*; Hahn, H.K., Mori, K., Eds.; SPIE: San Diego, CA, USA, 2019; p. 27.
19. Li, J.; Wang, W. Deployment and Application of Deep Learning Models under Computational Constraints. In Proceedings of the 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 15–18 December 2023; pp. 2529–2533.
20. Hung, C.-L.; Hsin, C.; Wang, H.-H.; Tang, C.Y. Optimization of GPU Memory Usage for Training Deep Neural Networks. In *Pervasive Systems, Algorithms and Networks*; Esposito, C., Hong, J., Choo, K.-K.R., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 289–293.

21. Iqbal, S.; Qureshi, A.N.; Li, J.; Choudhry, I.A.; Mahmood, T. Dynamic Learning for Imbalanced Data in Learning Chest X-Ray and CT Images. *Heliyon* **2023**, *9*, e16807. [CrossRef] [PubMed]
22. Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *J. Healthc. Eng.* **2019**, *2019*, 4180949. [CrossRef]
23. Liz, H.; Sánchez-Montañés, M.; Tagarro, A.; Domínguez-Rodríguez, S.; Dagan, R.; Camacho, D. Ensembles of Convolutional Neural Network Models for Pediatric Pneumonia Diagnosis. *Future Gener. Comput. Syst.* **2021**, *122*, 220–233. [CrossRef]
24. Ha Pham, N.; Tran, G.S. Apply a CNN-Based Ensemble Model to Chest-X Ray Image-Based Pneumonia Classification. *J. Adv. Inf. Technol.* **2024**, *15*, 1205–1214. [CrossRef]
25. Kahwachi, W.T.; Saed, K.A. A Novel Architecture of Convolutional Neural Network to Diagnose COVID-19 Disease. *Math. Stat. Eng. Appl.* **2022**, *71*, 4831–4856. Available online: <https://www.philstat.org/index.php/MSEA/article/view/1080> (accessed on 6 February 2025).
26. Luján-García, J.E.; Yáñez-Márquez, C.; Villuendas-Rey, Y.; Camacho-Nieto, O. A Transfer Learning Method for Pneumonia Classification and Visualization. *Appl. Sci.* **2020**, *10*, 2908. [CrossRef]
27. Jain, R.; Nagrath, P.; Kataria, G.; Sirish Kaushik, V.; Jude Hemanth, D. Pneumonia Detection in Chest X-Ray Images Using Convolutional Neural Networks and Transfer Learning. *Measurement* **2020**, *165*, 108046. [CrossRef]
28. Kaya, M. Feature Fusion-Based Ensemble CNN Learning Optimization for Automated Detection of Pediatric Pneumonia. *Biomed. Signal Process. Control* **2024**, *87*, 105472. [CrossRef]
29. Wang, K.; Jiang, P.; Meng, J.; Jiang, X. Attention-Based DenseNet for Pneumonia Classification. *IRBM* **2022**, *43*, 479–485. [CrossRef]
30. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131.e9. [CrossRef] [PubMed]
31. Keras 3 API Documentation. Available online: <https://keras.io/api/> (accessed on 24 September 2024).
32. Module: Tf | TensorFlow v2.16.1. Available online: https://www.tensorflow.org/api_docs/python/tf (accessed on 4 May 2024).
33. Sun, Y.; Zhang, Y.; Liu, S.; Lu, W.; Li, X. Image Super-Resolution Using Supervised Multi-Scale Feature Extraction Network. *Multimed. Tools Appl.* **2021**, *80*, 1995–2008. [CrossRef]
34. Shibu, D.S.; Priyadharsini, S.S. Multi Scale Decomposition Based Medical Image Fusion Using Convolutional Neural Network and Sparse Representation. *Biomed. Signal Process. Control* **2021**, *69*, 102789. [CrossRef]
35. Liu, J.; Zhang, L.; Guo, A.; Gao, Y.; Zheng, Y. Multi-Scale Feature Fusion Convolutional Neural Network for Multi-Modal Medical Image Fusion. In Proceedings of the 2023 4th International Conference on Computing, Networks and Internet of Things, Xiamen, China, 26–28 May 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 913–917.
36. Haque, M.I.U.; Dubey, A.K.; Danciu, I.; Justice, A.C.; Ovchinnikova, O.S.; Hinkle, J.D. Effect of Image Resolution on Automated Classification of Chest X-Rays. *J. Med. Imaging Bellingham Wash* **2023**, *10*, 044503. [CrossRef]
37. Baltruschat, I.M.; Nickisch, H.; Grass, M.; Knopp, T.; Saalbach, A. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. *Sci. Rep.* **2019**, *9*, 6381. [CrossRef]
38. Traore, B.B.; Kamsu-Foguem, B.; Tangara, F. Deep Convolution Neural Network for Image Recognition. *Ecol. Inform.* **2018**, *48*, 257–268. [CrossRef]
39. Misra, D. Mish: A Self Regularized Non-Monotonic Activation Function. *arXiv* **2020**, arXiv:1908.08681.
40. Radočaj, P.; Radočaj, D.; Martinović, G. Image-Based Leaf Disease Recognition Using Transfer Deep Learning with a Novel Versatile Optimization Module. *Big Data Cogn. Comput.* **2024**, *8*, 52. [CrossRef]
41. Radočaj, P.; Radočaj, D.; Martinović, G. Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays. *Big Data Cogn. Comput.* **2025**, *9*, 25. [CrossRef]
42. Vrbaničič, G.; Podgorelec, V. Efficient Ensemble for Image-Based Identification of Pneumonia Utilizing Deep CNN and SGD with Warm Restarts. *Expert Syst. Appl.* **2022**, *187*, 115834. [CrossRef]
43. Jaganathan, D.; Balsubramaniam, S.; Sureshkumar, V.; Dhanasekaran, S. Concatenated Modified LeNet Approach for Classifying Pneumonia Images. *J. Pers. Med.* **2024**, *14*, 328. [CrossRef] [PubMed]
44. Mabrouk, A.; Díaz Redondo, R.P.; Dahou, A.; Abd Elaziz, M.; Kayed, M. Pneumonia Detection on Chest X-Ray Images Using Ensemble of Deep Convolutional Neural Networks. *Appl. Sci.* **2022**, *12*, 6448. [CrossRef]
45. AlGhamdi, A.S. Efficient Deep Learning Approach for the Classification of Pneumonia in Infants from Chest X-Ray Images. *Trait. Signal* **2024**, *41*, 1245–1262. [CrossRef]
46. Mardianto, M.; Yoani, A.; Soewignjo, S.; Putra, I.; Dewi, D.A. Classification of Pneumonia from Chest X-Ray Images Using Support Vector Machine and Convolutional Neural Network. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 1015. [CrossRef]
47. Manickam, A.; Jiang, J.; Zhou, Y.; Sagar, A.; Soundrapandiyan, R.; Samuel, R. Automated Pneumonia Detection on Chest X-Ray Images: A Deep Learning Approach with Different Optimizers and Transfer Learning Architectures. *Measurement* **2021**, *184*, 109953. [CrossRef]

48. Dzhaynakbaev, N.; Kurmanbekkyzy, N.; Baimakhanova, A.; Mussatayeva, I. 2D-CNN Architecture for Accurate Classification of COVID-19 Related Pneumonia on X-Ray Images. *Int. J. Adv. Comput. Sci. Appl. IJACSA* **2024**, *15*, 905. [[CrossRef](#)]
49. Lan, X.; Zhang, Y.; Yuan, W.; Shi, F.; Guo, W. Image-Based Deep Learning in Diagnosing Mycoplasma Pneumonia on Pediatric Chest X-Rays. *BMC Pediatr.* **2024**, *24*, 720. [[CrossRef](#)] [[PubMed](#)]
50. Yi, R.; Tang, L.; Tian, Y.; Liu, J.; Wu, Z. Identification and Classification of Pneumonia Disease Using a Deep Learning-Based Intelligent Computational Framework. *Neural Comput. Appl.* **2023**, *35*, 14473–14486. [[CrossRef](#)]
51. Tejani, A.S.; Klontzas, M.E.; Gatti, A.A.; Mongan, J.T.; Moy, L.; Park, S.H.; Kahn, C.E., Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiol. Artif. Intell.* **2024**, *6*, e240300. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Interpretable Deep Learning for Pediatric Pneumonia Diagnosis Through Multi-Phase Feature Learning and Activation Patterns

Petra Radočaj ^{1,*} and Goran Martinović ² ¹ Layer d.o.o., Vukovarska Cesta 31, 31000 Osijek, Croatia² Faculty of Electrical Engineering, Computer Science and Information Technology, Josip Juraj Strossmayer University of Osijek, Kneza Trpimira 2B, 31000 Osijek, Croatia; goran.martinovic@ferit.hr

* Correspondence: pradocaj@layer.agency; Tel.: +385-31-224-600

Abstract: Pediatric pneumonia remains a critical global health challenge requiring accurate and interpretable diagnostic solutions. Although deep learning has shown potential for pneumonia recognition on chest X-ray images, gaps persist in understanding model interpretability and feature learning during training. We evaluated four convolutional neural network (CNN) architectures, i.e., InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2, using three approaches—standard convolution, multi-scale convolution, and strided convolution—all incorporating the Mish activation function. Among the tested models, InceptionResNetV2, with strided convolutions, demonstrated the best performance, achieving an accuracy of 0.9718. InceptionV3 also performed well using the same approach, with an accuracy of 0.9684. For DenseNet201 and MobileNetV2, the multi-scale convolution approach was more effective, with accuracies of 0.9676 and 0.9437, respectively. Gradient-weighted class activation mapping (Grad-CAM) visualizations provided critical insights, e.g., multi-scale convolutions identified diffuse viral pneumonia patterns across wider lung regions, while strided convolutions precisely highlighted localized bacterial consolidations, aligning with radiologists' diagnostic priorities. These findings establish the following architectural guidelines: strided convolutions are suited to deep hierarchical CNNs, while multi-scale approaches optimize compact models. This research significantly advances the development of interpretable, high-performance diagnostic systems for pediatric pneumonia using chest X-rays, bridging the gap between computational innovation and clinical application.

Keywords: pediatric pneumonia; convolutional neural networks; Mish activation function; multi-scale convolution; strided convolution; model interpretability; feature extraction



Academic Editor: Jian Sun

Received: 29 March 2025

Revised: 30 April 2025

Accepted: 6 May 2025

Published: 7 May 2025

Citation: Radočaj, P.; Martinović, G.

Interpretable Deep Learning for

Pediatric Pneumonia Diagnosis

Through Multi-Phase Feature

Learning and Activation Patterns.

Electronics **2025**, *14*, 1899. <https://doi.org/10.3390/electronics14091899>**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Worldwide morbidity and mortality from pediatric pneumonia persist as the key condition affecting children under five years old [1], with the highest burden observed in low- and middle-income countries (LMICs) [2]. Pneumonia causes 14% of deaths in children aged five and under, based on data from the World Health Organization, resulting in an estimated 740,000 fatalities annually [3]. Globally, the disease contributes to 10–20 million hospitalizations each year, with an incidence of 150–156 million cases, of which over 80% occur in LMICs [1,4]. Regions such as South Asia and Sub-Saharan Africa are disproportionately affected, with pneumonia being a leading cause of childhood mortality and a significant contributor to healthcare system strain [4]. Viral pathogens

remain the main cause of pneumonia, but the introduction of vaccines against *Streptococcus pneumoniae* and *Haemophilus influenzae* has led to significant decline in bacterial pneumonia cases [5–7]. Nevertheless, *S. pneumoniae* and *Mycoplasma pneumoniae* remain the leading bacterial pneumonia pathogens affecting vaccinated pediatric patients beyond the neonatal stage [8,9]. The symptoms of pediatric pneumonia clinically present as cough with fever and quick breathing, along with signs of respiratory problems, often accompanied by systemic manifestations such as fatigue, vomiting, and decreased appetite [10,11]. The severity of the condition can lead to complications involving pleural effusion, hypoxemia, and respiratory failure, necessitating prompt diagnosis and intervention [12,13]. Chest X-rays are crucial for diagnosing pediatric pneumonia by visualizing key lung pathologies such as consolidation, interstitial infiltrates, and pleural effusion. However, interpreting chest X-rays remains a challenging task for radiologists, especially when assessing early-stage infections or when clinical symptoms match those of other respiratory conditions [14,15]. Moreover, access to radiologists and advanced imaging technologies varies, along with the ability to perform feature extraction [16]. CNNs excel at identifying complex patterns in imaging data, making them particularly well-suited for tasks such as pneumonia recognition in chest X-rays. However, the “black box” nature of these models has historically limited their clinical adoption, as healthcare providers require interpretable and transparent decision-making processes to trust and effectively utilize AI-driven tools [17–19]. This challenge has spurred the development of visualization techniques, such as Grad-CAM, which provide insights into the regions of an image that influence a model’s predictions [20]. By generating heatmaps that highlight areas of interest, Grad-CAM enables clinicians to understand how a CNN arrives at its conclusions, thereby bridging the gap between AI and clinical practice [21]. The integration of Grad-CAM into deep learning frameworks for medical image analysis offers several advantages. First, it improves model transparency by visually delineating the most influential regions in an image, such as areas of consolidation or interstitial infiltrates in pneumonia cases. This interpretability fosters clinician trust and enables the identification of potential biases or errors, allowing for iterative model refinement [22,23]. Second, Grad-CAM supports the validation of model predictions by ensuring alignment between highlighted regions and established clinical features, reinforcing the model’s focus on biologically relevant areas [24,25]. Lastly, these visualization techniques serve as valuable educational tools, aiding less experienced clinicians in recognizing the subtle radiographic manifestations of pneumonia that might otherwise be overlooked [26].

Despite significant advancements in deep learning for medical image analysis, the application of deep learning visualization techniques to pneumonia classification, particularly in pediatric cases, remains underexplored. A comparative analysis of research trends from 2015 to 2024 indicates a substantial increase in studies focused on deep learning visualization in medicine, reaching a peak of approximately 450 publications in 2021. While the Web of Science Core Collection [27] indexes a substantial volume of scientific publications, research specifically focused on pneumonia represents a comparatively minor subset, peaking at 22 publications in 2021 and subsequently exhibiting a decline, as quantitatively depicted in Figure 1. This trend underscores a critical research gap in the application of deep learning visualization for pneumonia classification, particularly in pediatric populations. While established visualization methodologies such as Grad-CAM have been integrated into emerging research to enhance pneumonia recognition using CNNs, the interpretability of CNN models in this domain remains insufficiently investigated. Key areas requiring further exploration include feature activation patterns and the optimization of training efficiency.

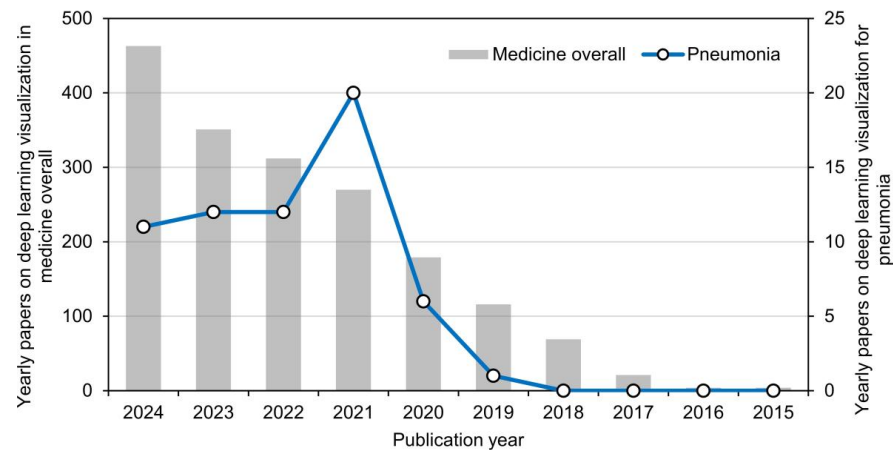


Figure 1. Comparative distribution of papers indexed in the Web of Science Core Collection showing machine learning visualization applications in pneumonia and medicine in general from 2015 to 2024.

Several research gaps persist in the application of deep learning for pediatric pneumonia classification. First, there is limited understanding of how CNNs process and interpret images during various phases of training [22,26]. Identifying the stages of training that most significantly contribute to learning and analyzing the evolution of feature activation over time could provide valuable insights for optimizing model performance and interpretability. This is particularly relevant for pediatric pneumonia, where subtle radiographic patterns necessitate precise feature extraction and analysis. Second, while advanced activation functions such as Mish have demonstrated superior results regarding smoothness, as well as improved gradient flow capabilities compared to those of traditional functions like ReLU [28], their potential in pediatric pneumonia diagnosis remains largely untapped. The ability of Mish to recognize complex features within chest radiographs has not been fully leveraged, revealing opportunities for improvements in diagnostic accuracy and computational efficiency. Third, although visualization techniques like Grad-CAM have been applied to medical imaging, their use in analyzing feature activation patterns across diverse CNN architectures—such as standard CNNs, multi-scale CNNs, and strided CNNs—has not been systematically explored. Understanding how these architectures differ in their capacity to highlight clinically relevant features in pediatric pneumonia cases is essential for developing more interpretable and reliable models. Finally, there is a pressing need for research that balances diagnostic accuracy with computational efficiency, particularly for deployment in resource-constrained settings where pediatric pneumonia is most prevalent. Many existing models are computationally intensive, limiting their applicability in low-resource environments [29]. Addressing these gaps is critical for advancing the field of pediatric pneumonia diagnosis and ensuring the development of effective and accessible AI-driven tools. This research provides valuable additions to pediatric pneumonia diagnostic methods and deep learning in the following ways:

- It provides a detailed analysis of how CNNs perceive and process images during different training phases, identifying critical learning stages that enhance model performance and interpretability in pediatric pneumonia classification.
- By evaluating three distinct CNN architectures—standard, multi-scale, and strided—this research offers a comprehensive comparison of their strengths and limitations, guiding the selection of optimal models for pediatric pneumonia diagnosis.
- Leveraging the Mish activation function and Grad-CAM visualization, this work enhances model transparency and diagnostic accuracy, enabling clinicians to better understand and trust AI-driven tools for pediatric pneumonia recognition.

Following this introduction, the paper proceeds as follows: Section 2 reviews the relevant literature. Section 3 provides a detailed description of the models and methodologies used. Section 4 presents and analyzes the key results. Finally, Section 5 concludes the paper with summary remarks and suggestions for future work.

2. Related Works

Advances in deep learning, particularly CNNs, have substantially enhanced the accuracy and interpretability of medical image analysis, particularly in the context of pediatric pneumonia diagnosis.

Luján-García et al. [30] utilized deep learning for pneumonia diagnosis in children under five, employing transfer learning with the pre-trained Xception network to classify 3883 pneumonia and 1349 normal chest X-ray images. The model achieved a precision of 0.84, a recall of 0.99, an F1 score of 0.91, and an AUC of 0.97, demonstrating competitive performance. They applied Grad-CAM to generate heatmaps, enabling the localization of pneumonia-related abnormalities. The study underscores the efficacy of transfer learning and visualization techniques in enhancing deep learning-based medical image analysis. Panwar et al. [31] developed a deep transfer learning model to enhance the accuracy and interpretability of COVID-19 detection using chest X-ray and CT-scan images. They integrated Grad-CAM for infection localization and implemented an early stopping mechanism to mitigate overfitting. The model achieved accuracies of 0.9655 in COVID-19 detection, 0.9404 in distinguishing COVID-19 from non-COVID cases, and 0.8947 in differentiating COVID-19 from normal cases, demonstrating the efficacy of deep learning in reliable and interpretable COVID-19 diagnosis. Zebin and Rezvy [32] developed a transfer learning pipeline for automated COVID-19 classification using chest X-ray images from publicly available datasets. They employed multiple pre-trained convolutional backbones as feature extractors, achieving classification accuracies of 0.90 with VGG16, 0.94 with ResNet50, and 0.97 with EfficientNetB0. To address data imbalance, they trained a CycleGAN for the synthetic augmentation of COVID-19 cases. Additionally, they implemented Grad-CAM to enhance model interpretability, enabling the visualization of affected lung regions for diagnosis and monitoring disease progression. Rahman et al. [33] proposed a VGG-16-based deep learning framework for explainable COVID-19 and pneumonia classification using chest X-ray images. They incorporated image enhancement, ROI segmentation, and data augmentation to improve accuracy. Additionally, they introduced a multi-layer gradient-weighted class activation mapping (ML-Grad-CAM) algorithm to generate class-specific saliency maps and a severity assessment index (SAI) to quantify infection severity. Their model achieved an accuracy of 0.9644 in a three-class classification task, demonstrating the potential of saliency maps for both diagnostic interpretation and severity assessment. Mohagheghi et al. [34] proposed two methods for COVID-19 diagnosis and differentiation from viral pneumonia using X-ray images. They employed deep neural networks for classification and an image retrieval approach for discrimination, both trained on healthy, pneumonia, and COVID-19 cases. Transfer learning and hashing functions enhanced the performance, achieving an accuracy of 0.97 for CNN-based classification and an overall precision of 0.87 for retrieval. Additionally, they introduced a decision support system integrating image retrieval and visualization techniques, including CT involvement score calculation, to provide physicians with interpretable diagnostic insights. Owais et al. [35] developed a lightweight deep learning ensemble for COVID-19 diagnosis using CT-scan and X-ray images, incorporating MobileNet, ShuffleNet, and FCNet to reduce the trainable parameters to 3.16 million. They introduced a multilevel class activation mapping (ML-CAM) layer to enhance lesion visualization, facilitating radiologist-assisted validation. A novel hierarchical training procedure dynamically adjusted epochs based on validation

performance, optimizing model convergence. The proposed model achieved F1-scores of 0.9460 (CT) and 0.9594 (X-ray), with AUCs of 0.9750 and 0.9799, respectively, demonstrating outstanding diagnostic accuracy and computational efficiency.

While these studies point to the significant capabilities of deep learning in medical image analysis, there remains a need to investigate how CNNs perceive images during different training phases, particularly in the context of pediatric pneumonia classification. Understanding which training stages contribute the most to learning and how feature activation evolves over time can provide valuable insights into optimizing model performance and interpretability. This study aims to address these gaps by analyzing feature activation maps using Grad-CAM and examining loss convergence across various CNN architectures, including a standard CNN, a CNN with multi-scale convolution, and a CNN with strided convolution. In each of these approaches, we employ the Mish activation function, which has demonstrated superior performance in terms of smoothness and gradient flow compared to that of traditional activation functions like ReLU [28,36,37]. By leveraging Mish, we aim to enhance the model's ability to capture complex patterns in chest X-ray images, particularly in the context of pediatric pneumonia. Furthermore, by identifying critical learning phases and optimizing training efficiency, this paper aims to improve the interpretability of deep learning models while increasing their diagnostic accuracy when classifying pediatric pneumonia.

3. Materials and Methods

In this study, we introduce an interpretable deep learning framework for pediatric pneumonia diagnosis, integrating multi-phase feature learning and activation pattern analysis. The methodology involves three stages, as presented in Figure 2: (1) data preparation and image preprocessing, including categorization into healthy and pneumonia classes, using data augmentation techniques; (2) implementation and evaluation of InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2 architectures, where three convolutional approaches were investigated: Approach 1, standard convolutions; Approach 2, multi-scale convolutions; and Approach 3, strided convolutions, each combined with Mish activation; and (3) performance assessment using standard metrics, with interpretability analysis via Grad-CAM.

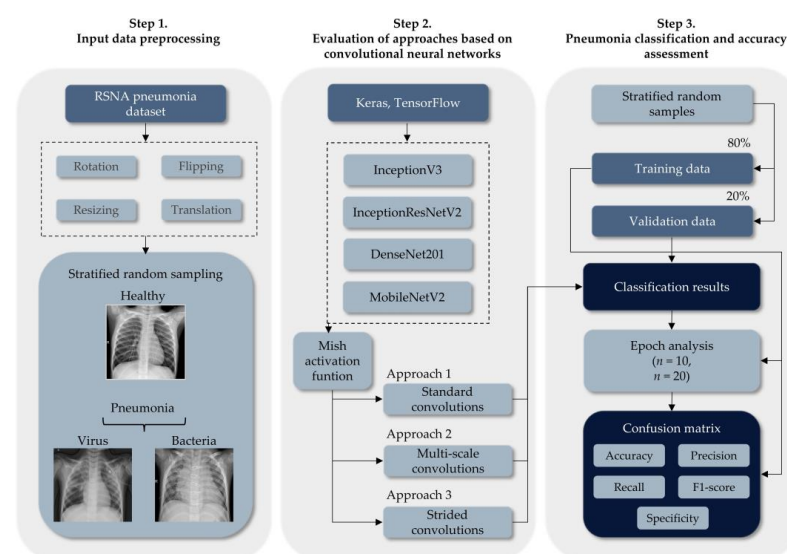


Figure 2. Interpretable pediatric pneumonia diagnosis methodology: (1) data preparation and augmentation; (2) evaluation of four CNNs, InceptionV3, InceptionResNetV2, DenseNet201, MobileNetV2, with three convolutional approaches—standard, multi-scale, and strided convolutions—with Mish activation function; (3) performance and Grad-CAM interpretability.

3.1. Data Preprocessing and Experimental Setup

We employed the publicly available Chest X-Ray Images (Pneumonia) dataset [38], which comprises 5856 pediatric chest X-ray scans collected from patients aged one to five years at Guangzhou Women and Children's Medical Center. The dataset consists of 4273 pneumonia cases and 1583 healthy controls [38]. To ensure data quality, an initial screening of all chest radiographs was conducted to eliminate low-quality or unreadable scans before analysis. Diagnosis assessment was conducted by two expert physicians who graded the images. A third expert validated the evaluation dataset to ensure accuracy and improve reliability. We partitioned the dataset using a stratified random split method at an 80:20 ratio, preserving the inherent class distribution. Although the dataset is imbalanced, we adopted multiple strategies to mitigate the impact of this characteristic. First, we applied extensive data augmentation—including rotation, translation, scaling, and horizontal flipping—equally across both classes to artificially expand the training set and reduce overfitting, specifically benefiting the minority (healthy) class. Second, we focused on evaluation metrics sensitive to class imbalance, such as precision, recall, and F1-score, rather than relying solely on accuracy. These metrics provided a more reliable assessment of model performance across both pneumonia and healthy classes. Third, stratified splitting ensured that the imbalance was consistently represented in both the training and validation sets, preventing bias toward the majority class during evaluation. We opted for a single split over k-fold cross-validation due to computational limitations, optimizing processing efficiency while maintaining comparable model performance. To standardize input dimensions, we resized all images to 224×224 pixels.

We built and trained the deep learning models in a Python 3.10 (Python Software Foundation, Wilmington, DE, USA) environment in Google Colab, leveraging the Keras-GPU [39] and TensorFlow-GPU [40] frameworks. We expedited training using NVIDIA Tesla K80 GPUs (Nvidia Corporation, Santa Clara, CA, USA) with 12 GB of memory. For each model, we set the training regimen to 20 epochs with a batch size of 32 and used the Adam optimizer for dynamic parameter updates. To prevent overfitting and improve convergence, we implemented an adaptive learning rate adjustment mechanism, reducing the rate upon validation performance stagnation, with a lower bound of 0.5×10^{-6} .

3.2. Interpretability and Convolutional Methods in Regards to Pneumonia Recognition

In this study, we employed four state-of-the-art CNN architectures for pneumonia classification from pediatric chest X-ray images: InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2. Each architecture presents unique characteristics that influence feature extraction, model complexity, and interpretability, particularly in the medical imaging domain, as demonstrated in Table 1.

InceptionV3, for instance, utilizes factorized convolutions and multi-scale processing via its Inception modules, enabling efficient, parallel feature extraction. It is particularly beneficial for capturing diverse patterns indicative of pulmonary infections and balances network depth against computational cost, a relevant consideration for pediatric datasets potentially exhibiting subtle signs [41–43]. InceptionResNetV2 enhances this ability by merging Inception's multi-scale approach with residual connections, improving gradient flow to stabilize the training of very deep networks. This architecture is well-suited for learning complex, hierarchical features from detailed medical images, such as pediatric radiographs, while maintaining training robustness [44]. DenseNet201 adopts a dense connectivity pattern in which each layer accesses feature maps from all the preceding layers, promoting extensive feature reuse and efficient information propagation. The integration of multi-level features and its inherent parameter efficiency make it advantageous for tasks like pneumonia recognition, especially with limited datasets [45,46]. Finally, MobileNetV2

prioritizes computational efficiency through depth-wise separable convolutions and an inverted residual structure, with linear bottlenecks. This design substantially reduces parameters and floating-point operations, rendering it highly suitable for resource-constrained environments, although its streamlined architecture may offer less sensitivity to extremely subtle features compared to that provided by more complex models [43,47].

Table 1. Comparative characteristics of CNN models employed for medical imaging.

Architecture	Strengths	Limitations
InceptionV3	Efficient multi-scale feature extraction through Inception modules [41]; effective at capturing fine-grained patterns, with moderate computational cost [42].	May struggle with very subtle features in pediatric lungs; not as lightweight as MobileNetV2 [43].
InceptionResNetV2	Combines Inception modules with residual connections; deeper and more accurate; improved gradient flow [44].	Higher computational requirements; risk of overfitting, if not carefully regularized [44].
DenseNet201	Dense connectivity promotes feature reuse and mitigates vanishing gradients [45]; strong performance on small datasets.	Higher memory usage; feature maps can become redundant, slightly increasing inference time [37,46].
MobileNetV2	Lightweight with depth-wise separable convolutions; ideal for real-time applications and devices with limited resources [43].	May underperform on very complex patterns when compared to the results for heavier models; limited representational capacity [47].

Grad-CAM is a visualization technique used to interpret the predictions of CNNs by highlighting important regions in an input image [20,31]. This method utilizes the gradient information flowing into the final convolutional layer to produce a coarse localization map of the salient regions. By visualizing these areas, Grad-CAM aids in the interpretability of deep learning models, which is crucial in medical applications such as pneumonia classification, where understanding why a model makes a particular decision can improve trust and reliability [48]. The class-discriminative localization map L^c for a target class is computed as defined in Equation (1):

$$L^c = \text{ReLU} \sum_k \alpha_k^c A^k \quad (1)$$

where A^k represents the activation maps of the last convolutional layer, and α_k^c represents importance weights computed as defined in Equation (2):

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (2)$$

Here, Z is the number of spatial locations, and y^c represents the class score for class c . The ReLU function ensures that only positive influences are considered, emphasizing relevant regions in the image that contribute to the model's decision [20,49]. This is particularly beneficial in pneumonia classification, where highlighting infected lung regions in chest X-ray images can assist radiologists in making more accurate diagnoses. Applying Grad-CAM in this study not only enhances model transparency but also allows for correlation between the architectural depth or complexity and the quality of the generated heatmaps. Models like DenseNet201 and InceptionResNetV2, due to their richer feature hierarchies, typically produce more distinct and clinically meaningful visual explanations compared to lightweight models like MobileNetV2.

In addition to interpretability techniques, activation functions play a crucial role in deep learning models. The Mish activation function, a non-monotonic function, has been shown to improve gradient flow and smoothness during training, leading to better feature extraction and generalization [36,37]. It is defined mathematically using Equation (3):

$$\text{Mish}(x) = x * \tanh(\text{softplus}(x)) = x * \tanh(\ln(1 + e^x)) \quad (3)$$

In contrast to ReLU, which sets the negative values to zero, Mish preserves small negative values and provides a smooth transition, which is particularly beneficial in medical image analysis, as demonstrated in recent studies. The smooth non-linearity helps in preserving finer details in pneumonia classification tasks, allowing for improved model robustness. Mish has also been observed to provide better feature representation compared to that of traditional activation functions, leading to more stable training and enhanced classification accuracy in deep learning architectures [36].

Deep learning models for pneumonia classification leverage different types of convolutional operations to extract meaningful features from chest X-ray images. Standard convolutions are fundamental for capturing spatial features by applying fixed-size kernels to detect the texture and structural patterns associated with pneumonia-infected lungs [50]. Multi-scale convolutions employ filters of varying sizes to detect abnormalities at different resolutions, ensuring that subtle and large-scale pneumonia-related features are equally recognized [41,51]. This enhances the model's ability to generalize across diverse manifestations of the disease, such as varying opacity and lesion size in infected lungs. Additionally, multi-scale convolutional layers help capture both local fine-grained details and broader anatomical structures, which are essential for distinguishing pneumonia from other lung conditions [52]. Strided convolutions serve a dual purpose of providing both feature extraction and dimensionality reduction, reducing computational complexity while preserving critical spatial information [53,54]. Unlike max-pooling, strided convolutions ensure a more structured downsampling process, which can be advantageous in medical image processing where fine details are crucial for accurate diagnosis. By reducing the spatial resolution while maintaining relevant features, strided convolutions facilitate deeper architectures without excessive computational costs [55]. Furthermore, when combined with residual connections, they mitigate information loss, ensuring that vital pneumonia-specific features are retained through the network layers.

By integrating Grad-CAM for model interpretability, Mish activation for enhanced feature extraction, and a combination of standard, multi-scale, and strided convolutions, deep learning-based pneumonia classification systems can achieve higher accuracy and robustness. These methodologies contribute to improved diagnostic performance, making CNN-based models more reliable for medical applications. The ability to highlight affected lung regions, capture multi-scale features, and efficiently process medical images enhances the potential for the use of AI-assisted diagnostic tools in clinical settings, leading to better patient outcomes and more informed medical decision making.

3.3. Pediatric Pneumonia Accuracy Assessment

We evaluated the performance of deep learning models for pneumonia classification in chest radiographs. We analyzed training and validation accuracy, loss, and key classification metrics, deriving them from the confusion matrix. These key metrics—precision, recall, specificity, F1-score, and accuracy—as defined in Equations (4)–(8), allowed us to perform a comprehensive evaluation of the models' diagnostic capability, as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (6)$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (7)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

We classified the model's outcomes as follows: true positives (TP), where it correctly identified pneumonia cases; true negatives (TN), where it correctly identified non-pneumonia cases; false positives (FP), where it incorrectly classified non-pneumonia cases as pneumonia; and false negatives (FN), where it incorrectly classified pneumonia cases as non-pneumonia. We calculated specificity, which reflects the model's ability to correctly identify non-pneumonia cases, and recall (sensitivity), which reflects its ability to identify pneumonia cases. Radiographic features indicative of pneumonia, including increased pulmonary opacity, consolidation, and pleural effusion, served as primary discriminative criteria for the model's classification algorithm. However, the potential for overlapping radiographic manifestations with alternative pathologies, coupled with the susceptibility to imaging artifacts, introduced sources of diagnostic error, resulting in classification inaccuracies.

4. Results and Discussion

In this study, we evaluated the efficacy of four deep learning architectures—InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2—for pediatric pneumonia classification. We implemented three distinct convolutional approaches—Approach 1, base model with Mish activation function and standard convolution; Approach 2, base model with Mish activation function and multi-scale convolutions; Approach 3, base model with Mish activation function and strided convolutions. We assessed the performance of each architecture and approach using established metrics, including accuracy, precision, recall, F1-score, and specificity, providing a comprehensive evaluation of their classification capabilities. For InceptionResNetV2, Approach 3 yielded the best results, achieving the highest accuracy of 0.9718 and F1-score of 0.9634. This approach demonstrated its ability to efficiently capture complex features while maintaining a high precision of 0.9767 and a recall of 0.9519. The combination of strided convolutions and the Mish activation function likely enhanced feature extraction and generalization [56], making it the top-performing model overall. Similarly, for InceptionV3, Approach 3 achieved its best performance with an accuracy of 0.9684 and an F1-score of 0.9595. This approach likely improved its hierarchical feature extraction process and increased the specificity to 0.9211. The high precision and recall indicate that the model effectively minimized false positives and false negatives, respectively [57]. In contrast, DenseNet201 performed optimally with Approach 2, achieving an accuracy of 0.9676 and an F1-score 0.9582. The multi-scale approach complemented the dense connectivity of DenseNet201 by capturing features at multiple resolutions, leading to a robust balance between sensitivity, with a recall of 0.9510, and specificity, with a score of 0.9148. MobileNetV2 also benefited from Approach 2, achieving an accuracy of 0.9437 and an F1-score of 0.9254. Despite its smaller size, this approach allowed it to efficiently extract diverse features [47], although its specificity of 0.8297 was lower compared to that of other models. Key observations highlight that Approach 3 was particularly effective for InceptionV3 and InceptionResNetV2, aligning well with their hierarchical feature extraction processes and improving their computational efficiency. Approach 2 worked best for DenseNet201 and MobileNetV2, enhancing their ability to capture features at multiple resolutions while maintaining competitive performance. The precision and F1-score metrics

revealed that all models effectively addressed class imbalance, with InceptionResNetV2 and DenseNet201 achieving the best balance between sensitivity and specificity. These findings underscore the importance of tailoring architectural modifications to the strengths of each model, as demonstrated in Table 2.

Table 2. Pneumonia classification accuracy assessment of proposed approaches.

Transfer Deep Learning Model	Classification Approach	Accuracy	F1-Score	Precision	Recall	Specificity
InceptionV3	Approach 1	0.9573	0.9462	0.9444	0.9479	0.9274
	Approach 2	0.9394	0.9202	0.9394	0.9049	0.8297
	Approach 3	0.9684	0.9595	0.9659	0.9536	0.9211
InceptionResNetV2	Approach 1	0.9684	0.9604	0.9564	0.9645	0.9558
	Approach 2	0.9539	0.9407	0.9483	0.9337	0.8896
	Approach 3	0.9718	0.9634	0.9767	0.9519	0.9085
MobileNetV2	Approach 1	0.9206	0.9060	0.8871	0.9367	0.9716
	Approach 2	0.9437	0.9254	0.9481	0.9078	0.8297
	Approach 3	0.9104	0.8945	0.8754	0.9277	0.9653
DenseNet201	Approach 1	0.9650	0.9542	0.9709	0.9403	0.8864
	Approach 2	0.9676	0.9582	0.9662	0.9510	0.9148
	Approach 3	0.9573	0.9440	0.9622	0.9291	0.8675

Approach 1, base model with Mish activation function and standard convolution; Approach 2, base model with Mish activation function and multi-scale convolutions; Approach 3, base model with Mish activation function and strided convolutions. The highest assessment metrics are bolded.

The confusion matrix provides a detailed breakdown of model performance across different classification approaches, as presented in Figure 3. For InceptionV3, Approach 3 yielded the best results, achieving a high true positive rate of 0.9860, with only 12 false negative cases. In contrast, Approach 2 struggled to distinguish healthy cases, leading to an increase in false positives. InceptionResNetV2 demonstrated consistent performance across all approaches, with Approach 3 producing the lowest false negative rate of just four cases and a recall of 0.9953, confirming its superior sensitivity in classifying pneumonia. MobileNetV2, however, exhibited higher false positive rates using Approach 3, indicating a tendency to misclassify healthy cases as pneumonia. The highest recall of 0.9860 was observed in Approach 2, demonstrating strong pneumonia detection capabilities but at the cost of increased false positives. DenseNet201 performed optimally with Approach 2, achieving a high specificity of 0.9148 and a recall of 0.9871 for pneumonia detection. However, Approach 3 resulted in a higher number of false positives, suggesting a trade-off in regards to specificity. Overall, Approach 3 proved the most effective for deeper models like InceptionV3 and InceptionResNetV2, while multi-scale convolutions in Approach 2 were more beneficial for MobileNetV2 and DenseNet201 [37,46,58]. This pattern aligns with the observed accuracy and F1-score trends, reinforcing the idea that different architectural enhancements impact classification trade-offs in unique ways.

The results, evaluated in terms of training accuracy, validation accuracy, and validation loss at 10 and 20 epochs, provide key insights into the learning dynamics and generalization capabilities of each model, as demonstrated in Table 3. Notably, Approach 3 demonstrated superior efficacy for InceptionV3 and InceptionResNetV2. InceptionV3 achieved a validation accuracy of 0.9583 at 10 epochs and 0.9705 at 20 epochs, with the lowest validation loss values of 0.1155 and 0.0744, respectively. Similarly, InceptionResNetV2 exhibited substantial improvement in validation accuracy from 0.9392 to 0.9670, while validation loss markedly decreased from 0.4888 to 0.0912. Conversely, Approach 2 yielded the most favorable results for DenseNet201 and MobileNetV2. DenseNet201

attained a validation accuracy of 0.9592 at 10 epochs and 0.9566 at 20 epochs, accompanied by validation loss values of 0.1171 and 0.1058, respectively. MobileNetV2 demonstrated the fastest convergence, achieving a validation accuracy of 0.9549 by 10 epochs, with a minimal change to 0.9523 at 20 epochs. Despite slightly higher validation loss values of 0.1289 at 10 epochs and 0.1383 at 20 epochs, MobileNetV2 exhibited stable training accuracy, recording 0.9688 at 10 epochs and 0.9520 at 20 epochs.

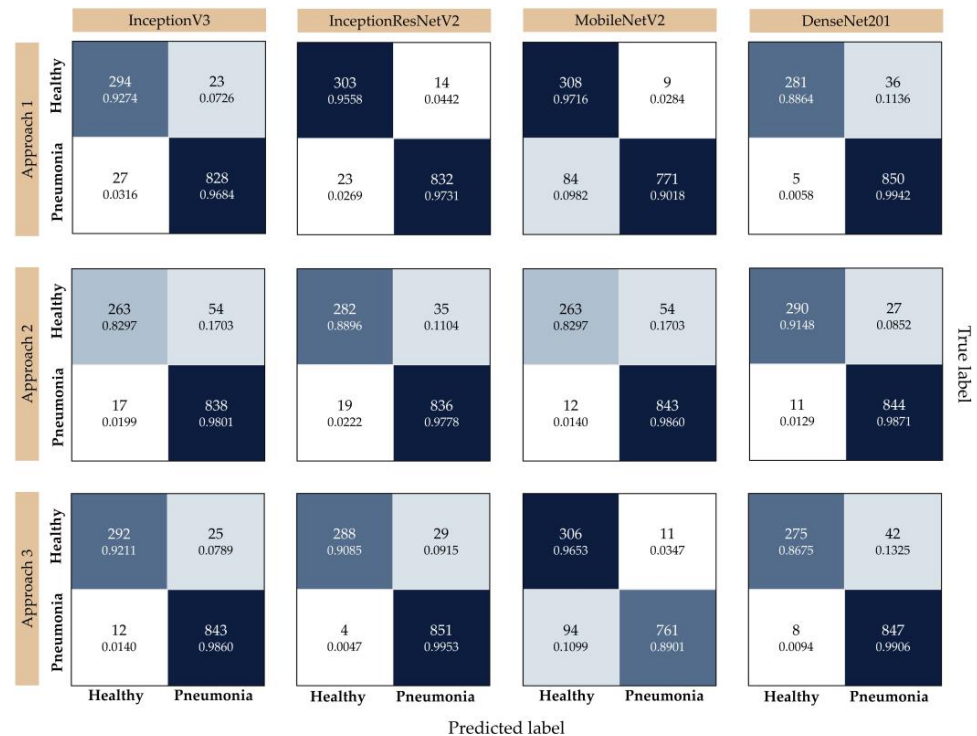


Figure 3. The confusion matrices for the proposed approaches. Within each cell, the count of classified instances is presented as the top number, and the corresponding percentage for that actual class is displayed as the bottom number.

Table 3. Accuracy and validation loss during 10 and 20 epochs of training for the proposed approaches.

Transfer Deep Learning Model	Classification Approach	10 Epochs			20 Epochs		
		TA	VA	VL	TA	VA	VL
InceptionV3	Approach 1	0.9379	0.9149	0.2379	0.9662	0.9714	0.1034
	Approach 2	0.9364	0.9253	0.1844	0.9688	0.9384	0.1645
	Approach 3	0.9638	0.9583	0.1155	0.9658	0.9705	0.0744
InceptionResNetV2	Approach 1	0.9688	0.9453	0.7087	0.9375	0.9714	0.0723
	Approach 2	0.9302	0.9392	0.1664	0.9688	0.9497	0.1476
	Approach 3	0.9512	0.9392	0.4888	0.9769	0.9670	0.0912
MobileNetV2	Approach 1	0.9375	0.9201	0.2278	0.9705	0.9193	0.2688
	Approach 2	0.9688	0.9549	0.1289	0.9520	0.9523	0.1383
	Approach 3	0.9664	0.8811	0.5385	0.9681	0.9314	0.2727
DenseNet201	Approach 1	0.9062	0.9314	0.2257	0.9688	0.9635	0.1372
	Approach 2	0.9688	0.9592	0.1171	0.9375	0.9566	0.1058
	Approach 3	0.9523	0.9583	0.1191	0.9062	0.9566	0.1254

Approach 1, base model with Mish activation function and standard convolution; Approach 2, base model with Mish activation function and multi-scale convolutions; Approach 3, base model with Mish activation function and strided convolutions; (TA) training accuracy; (VA) validation accuracy; (VL) validation loss. The highest assessment metrics are bolded.

From a generalization perspective, InceptionV3 achieved the lowest validation loss of 0.0744 at 20 epochs, closely followed by DenseNet201 at 0.1058. Overall, the results underscore the importance of selecting architectural modifications based on model depth and computational constraints. Approach 3 proved beneficial for deeper architectures such as InceptionV3 and InceptionResNetV2, whereas Approach 2 exhibited optimized performance in more compact models like DenseNet201 and MobileNetV2.

Moreover, we analyzed Grad-CAM visualizations for deep learning models—InceptionV3, InceptionResNetV2, MobileNetV2, and DenseNet201—in classifying healthy lungs, viral pneumonia, and bacterial pneumonia from chest X-ray images, using multiple Grad-CAM approaches to highlight key decision-making areas, as presented in Figure 4, Figure 5, and Figure 6, respectively. In healthy cases, models like InceptionV3 and DenseNet201 show minimal activation, primarily along the ribcage or lung periphery, while viral pneumonia is characterized by diffuse, bilateral activation across both lungs, often including the heart region [59,60], as seen in MobileNetV2 and DenseNet201, which effectively capture interstitial changes. Bacterial pneumonia, in contrast, is identified by sharp, localized activation, typically within one lung lobe [61], with DenseNet201 excelling in its detection due to its highly focused heatmaps, while InceptionResNetV2 also well differentiates the sharply defined activations in specific lobes. Grad-CAM approaches, particularly Approach 3, provide the clearest visualizations, revealing sharp, focused bacterial pneumonia regions and broader, generalized viral patterns, while Approach 1 highlights initial lung field activations, and Approach 2 refines distinctions between pneumonia types. MobileNetV2 performs best for viral pneumonia due to its strong central lung and heart activation, DenseNet201 is the most accurate for bacterial pneumonia, with its distinct lobar focus, and InceptionResNetV2 offers a balanced performance for both. Key takeaways include minimal activation in healthy cases, diffuse patterns in viral pneumonia, and localized, sharp activations in bacterial pneumonia, with DenseNet201 emerging as the most precise model for distinguishing between the two pneumonia types [45,62]. Overall, Grad-CAM effectively illustrates how these models interpret pneumonia patterns, confirming DenseNet201’s superiority for bacterial pneumonia and MobileNetV2’s strength in viral pneumonia detection.

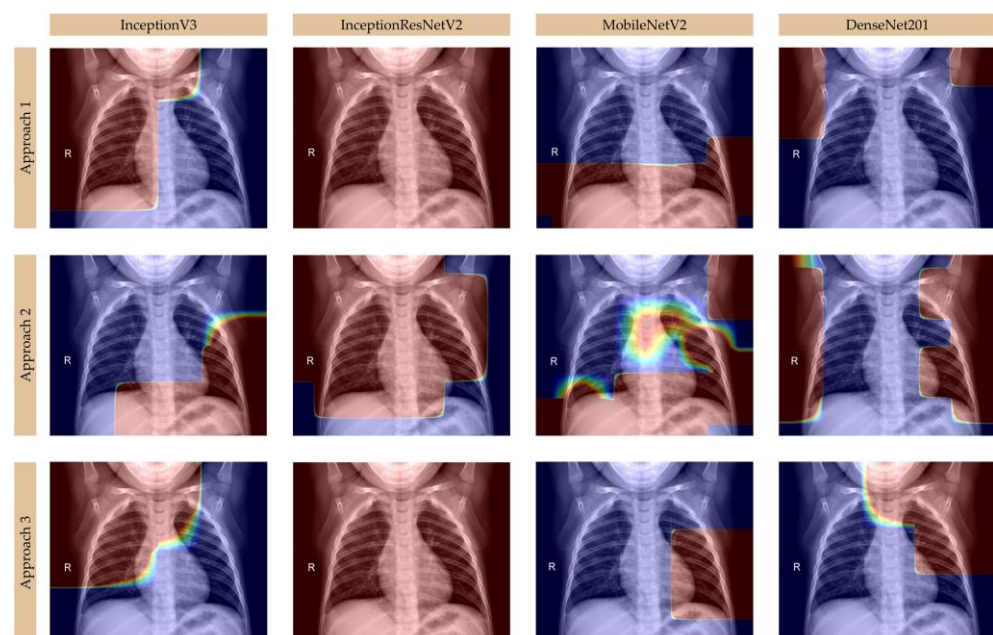


Figure 4. Grad-CAM visualizations for healthy lungs.

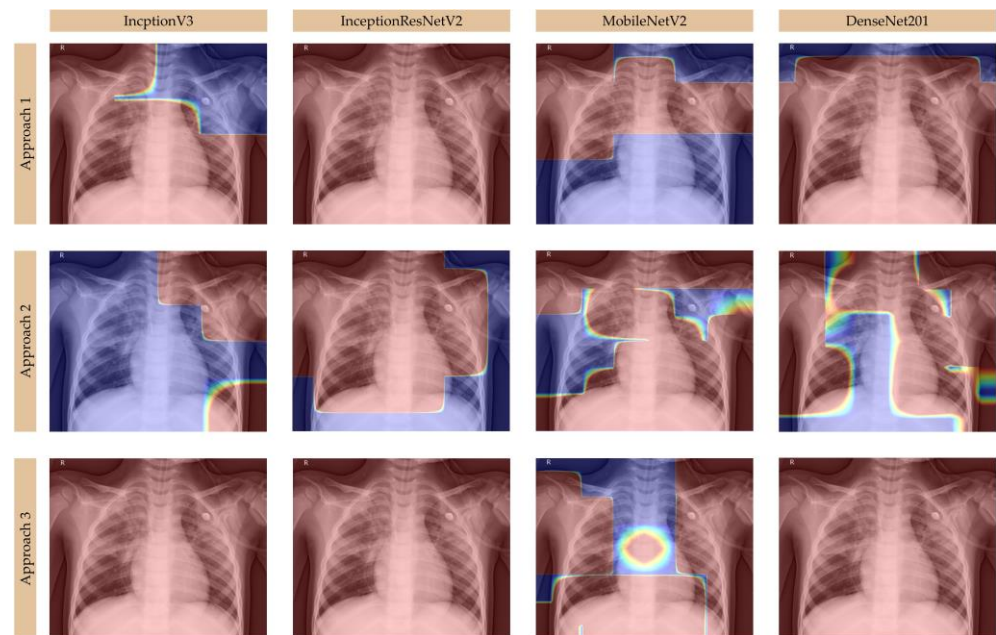


Figure 5. Grad-CAM visualizations for viral pneumonia.

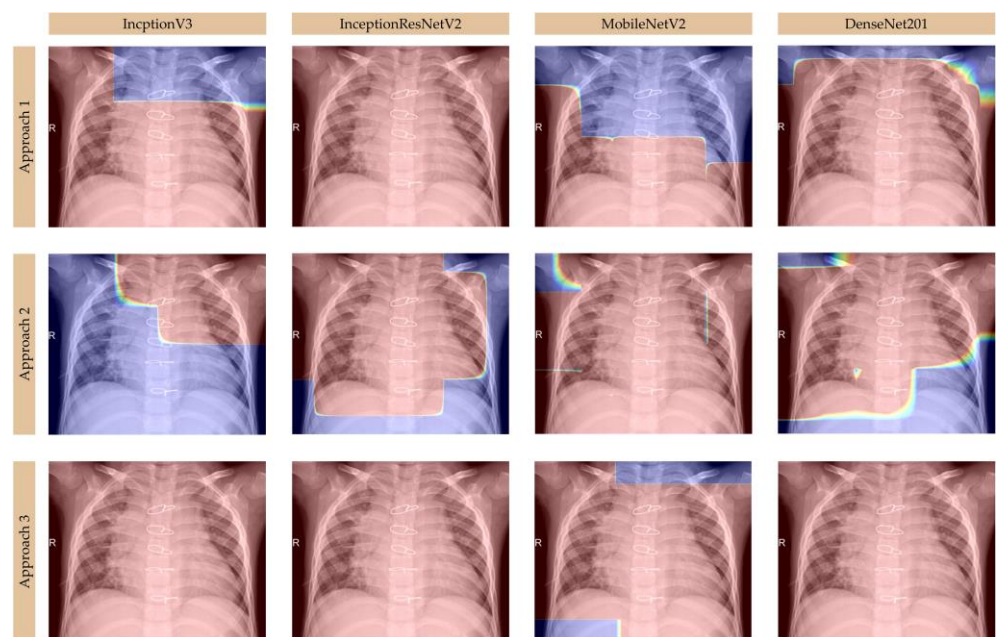


Figure 6. Grad-CAM visualizations for bacterial pneumonia.

The consistent performance of MobileNetV2 with Approach 2 across training and validation metrics, coupled with its rapid convergence and stability, made it a suitable choice for detailed interpretability analysis using Grad-CAM, as presented in Figure 7. Its ability to generalize well, with minimal overfitting, underscores the importance of selecting models that balance performance, efficiency, and interpretability, particularly in medical image classification tasks, where false positives and false negatives can have significant clinical implications [63,64].

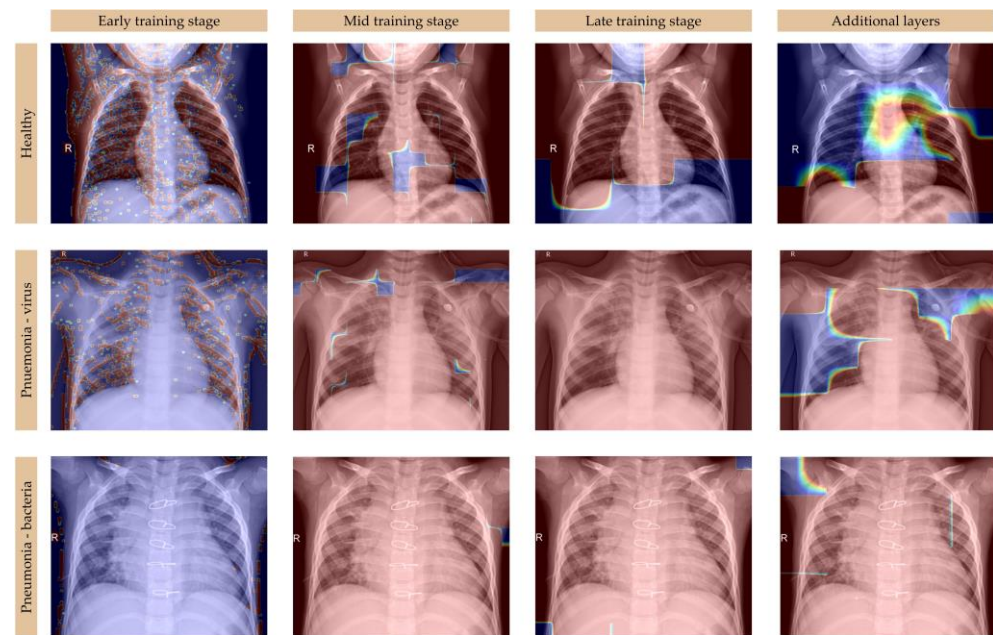


Figure 7. Grad-CAM visualizations of MobileNetV2 with Approach 2 across training stages for the classification of healthy lungs, viral pneumonia, and bacterial pneumonia.

The integration of multi-scale convolutions with Mish activation in MobileNetV2 significantly improves heatmap-based lung disease detection by refining feature extraction across multiple spatial resolutions. This architectural enhancement enables the model to simultaneously capture fine-grained abnormalities, such as the diffuse opacities characteristic of viral pneumonia, and larger consolidations typical of bacterial infections. The multi-scale approach effectively suppresses false activations in non-pathological regions, maintaining precise focus on diagnostically relevant areas throughout the training process [41,52]. During the initial training phases, the multi-scale architecture mitigates scattered attention patterns by promoting more meaningful feature extraction. As training progresses to intermediate stages, it enhances consistency in activation patterns across diverse lung pathologies. In the final training stages, this approach produces sharply defined heatmaps, while preserving sensitivity to subtle pathological indicators that might otherwise be overlooked. The Mish activation function offers superior gradient flow and feature diversity compared to those of conventional ReLU activation [28,36]. Its smooth, non-monotonic characteristics address several limitations of traditional activation functions. Specifically, Mish prevents vanishing gradients during backpropagation while maintaining richer feature representations. This property proves particularly valuable in medical image analysis, where subtle pathological patterns require precise detection [37]. In early training iterations, Mish activation helps avoid suboptimal initialization traps that can hinder model convergence. During the intermediate training phases, it strengthens the mid-level feature extraction capabilities. In the final training stages, Mish activation yields more confident and precise spatial activations in the generated heatmaps. The combined implementation of multi-scale convolutions and Mish activation provides multiple synergistic benefits. First, it produces heatmaps with superior pathological localization. Second, it enhances classification accuracy for distinguishing between viral and bacterial pneumonia manifestations. Third, it significantly reduces activation noise in healthy tissue regions. Fourth, it accelerates model convergence, potentially enabling early stopping strategies, without compromising diagnostic performance. This optimized architecture demonstrates particular value in regards to clinical decision support systems, where reliable detection of pulmonary abnormalities must be maintained across diverse imaging conditions and acquisition pro-

tocols. The improved interpretability of the resulting heatmaps provides clinicians with more trustworthy visual explanations of the model's diagnostic reasoning process.

This study has some limitations, primarily due to reliance on public datasets, which may introduce biases in regards to image quality and patient demographics, affecting generalizability. Future research should validate the model across diverse clinical settings to enhance robustness. Implementing a hybrid cloud-edge infrastructure could enable scalable deployment, balancing centralized model updates with local inference for improved data privacy. Additionally, long-term clinical integration requires continuous learning mechanisms. Incorporating radiologist feedback loops would allow the model to adapt to real-world cases, ensuring sustained diagnostic accuracy and practical clinical impact. Advanced deep learning approaches, particularly variational autoencoders (VAEs), have shown considerable value, especially those focused on analyzing COVID-19 chest X-rays. These methods have proven effective in addressing challenges like class imbalance and enhancing feature learning [65,66]. Future work could involve integrating VAE-based techniques for data balancing or feature extraction to further improve model robustness. Additionally, both highlighted sophisticated uses of Grad-CAM for model interpretability, suggesting that incorporating more advanced or modified Grad-CAM approaches could enhance explainability and clinical trust in pneumonia classification models.

5. Conclusions and Future Work

This study demonstrates that deep learning models can achieve excellent diagnostic performance for pediatric pneumonia when properly optimized. The InceptionResNetV2 model, with strided convolutions and Mish activation, achieved the highest accuracy of 0.9718, closely followed by InceptionV3 at 0.9684. The DenseNet201 architecture performed exceptionally well with multi-scale convolutions, reaching 0.9676 accuracy, while MobileNetV2 achieved 0.9437 accuracy using the same approach. These results highlight how different convolutional strategies can be matched to specific network architectures for optimal performance. The integration of Grad-CAM provided valuable interpretability, clearly visualizing diagnostic features in chest X-rays. For viral pneumonia, the models detected diffuse patterns across lung fields, while for bacterial cases, the models identified the precise localization of consolidations. This capability to explain decisions builds crucial trust for clinical adoption.

Moving forward, we aim to translate these research findings into clinical practice by developing an end-to-end AI diagnostic system capable of real-time pneumonia detection. This will require expanded validation across diverse patient populations and imaging protocols to ensure robustness. We also plan to optimize these models for practical deployment through the use of edge computing solutions that preserve diagnostic accuracy while meeting clinical latency requirements. Close collaboration with radiologists will be essential to refine model predictions and integrate AI assistance seamlessly into existing workflows. By bridging the gap between technical innovation and clinical needs, this work paves the way for more accurate, interpretable, and deployable AI tools for use in pediatric respiratory care.

Author Contributions: Conceptualization, P.R.; methodology, P.R.; software, P.R.; validation, G.M.; formal analysis, P.R.; investigation, P.R.; resources, P.R.; data curation, P.R.; writing—original draft preparation, P.R.; writing—review and editing, P.R. and G.M.; visualization, P.R.; supervision, G.M.; project administration, G.M.; funding acquisition, P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The code developed in this study is available on request from the corresponding author. The open access Chest X-Ray Images (Pneumonia) repository containing images collected from pediatric patients, is divided into two categories—Pneumonia/Healthy—and is available at <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia> (accessed on 15 March 2025).

Conflicts of Interest: Author Petra Radočaj was employed by the company Layer d.o.o. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Rudan, I.; O'Brien, K.L.; Nair, H.; Liu, L.; Theodoratou, E.; Qazi, S.; Lukšić, I.; Walker, C.L.F.; Black, R.E.; Campbell, H. Epidemiology and Etiology of Childhood Pneumonia in 2010: Estimates of Incidence, Severe Morbidity, Mortality, Underlying Risk Factors and Causative Pathogens for 192 Countries. *J. Glob. Health* **2013**, *3*, 010401. [PubMed]
2. Marangu, D.; Zar, H.J. Childhood Pneumonia in Low-and-Middle-Income Countries: An Update. *Paediatr. Respir. Rev.* **2019**, *32*, 3–9. [CrossRef] [PubMed]
3. Pneumonia in Children. Available online: <https://www.who.int/news-room/fact-sheets/detail/pneumonia> (accessed on 22 December 2024).
4. Genie, Y.D.; Sayih, A.; Dessalegn, N.; Adugnaw, E.; Hiwot, A.Y.; Tesfa, T.B.; Kindie, K.; Gutema, L.; Ayalew, E.; Kebede, B.F. Time to Recovery from Severe Pneumonia and Its Predictors among Pediatric Patients Admitted in South West Region Governmental Hospitals, South West Ethiopia: Prospective Follow-up Study. *Glob. Pediatr.* **2024**, *9*, 100227. [CrossRef]
5. Scotta, M.C.; Marostica, P.J.C.; Stein, R.T. 25—Pneumonia in Children. In *Kendig's Disorders of the Respiratory Tract in Children*, 9th ed.; Wilmott, R.W., Deterding, R., Li, A., Ratjen, F., Sly, P., Zar, H.J., Bush, A., Eds.; Elsevier: Philadelphia, PA, USA, 2019; pp. 427–438.e4; ISBN 978-0-323-44887-1.
6. Onwuchekwa, C.; Edem, B.; Williams, V.; Oga, E. Estimating the Impact of Pneumococcal Conjugate Vaccines on Childhood Pneumonia in Sub-Saharan Africa: A Systematic Review. *F1000Research* **2020**, *9*, 765. [CrossRef]
7. Pavia, M.; Bianco, A.; Nobile, C.G.A.; Marinelli, P.; Angelillo, I.F. Efficacy of Pneumococcal Vaccination in Children Younger than 24 Months: A Meta-Analysis. *Pediatrics* **2009**, *123*, e1103–e1110. [CrossRef]
8. de Groot, R.C.A.; Estevão, S.C.; Meyer Sauter, P.M.; Perkasa, A.; Hoogenboezem, T.; Spuesens, E.B.M.; Verhagen, L.M.; van Rossum, A.M.C.; Unger, W.W.J. Mycoplasma Pneumoniae Carriage Evades Induction of Protective Mucosal Antibodies. *Eur. Respir. J.* **2022**, *59*, 2100129. [CrossRef]
9. Berg, A.S.; Inchley, C.S.; Aase, A.; Fjaerli, H.O.; Bull, R.; Aaberge, I.; Leegaard, T.M.; Nakstad, B. Etiology of Pneumonia in a Pediatric Population with High Pneumococcal Vaccine Coverage: A Prospective Study. *Pediatr. Infect. Dis. J.* **2016**, *35*, e69–e75. [CrossRef]
10. Søndergaard, M.J.; Friis, M.B.; Hansen, D.S.; Jørgensen, I.M. Clinical Manifestations in Infants and Children with Mycoplasma Pneumoniae Infection. *PLoS ONE* **2018**, *13*, e0195288. [CrossRef]
11. Biagi, C.; Cavallo, A.; Rocca, A.; Pierantoni, L.; Antonazzo, D.; Dondi, A.; Gabrielli, L.; Lazzarotto, T.; Lanari, M. Pulmonary and Extrapulmonary Manifestations in Hospitalized Children with Mycoplasma Pneumoniae Infection. *Microorganisms* **2021**, *9*, 2553. [CrossRef]
12. de Benedictis, F.M.; Kerem, E.; Chang, A.B.; Colin, A.A.; Zar, H.J.; Bush, A. Complicated Pneumonia in Children. *Lancet* **2020**, *396*, 786–798. [CrossRef]
13. Zhang, J.; Zhu, Y.; Zhou, Y.; Gao, F.; Qiu, X.; Li, J.; Yuan, H.; Jin, W.; Lin, W. Pediatric Adenovirus Pneumonia: Clinical Practice and Current Treatment. *Front. Med.* **2023**, *10*, 1207568. [CrossRef] [PubMed]
14. Fancourt, N.; Deloria Knoll, M.; Barger-Kamate, B.; de Campo, J.; de Campo, M.; Diallo, M.; Ebruke, B.E.; Feikin, D.R.; Gleeson, F.; Gong, W.; et al. Standardized Interpretation of Chest Radiographs in Cases of Pediatric Pneumonia From the PERCH Study. *Clin. Infect. Dis.* **2017**, *64*, S253–S261. [CrossRef] [PubMed]
15. Salehi, M.; Mohammadi, R.; Ghaffari, H.; Sadighi, N.; Reiazi, R. Automated Detection of Pneumonia Cases Using Deep Transfer Learning with Paediatric Chest X-Ray Images. *Br. J. Radiol.* **2021**, *94*, 20201263. [CrossRef]
16. Chan, H.-P.; Samala, R.K.; Hadjiiski, L.M.; Zhou, C. Deep Learning in Medical Image Analysis. *Adv. Exp. Med. Biol.* **2020**, *1213*, 3–21. [CrossRef] [PubMed]
17. Hakkoum, H.; Abnane, I.; Idri, A. Interpretability in the Medical Field: A Systematic Mapping and Review Study. *Appl. Soft Comput.* **2022**, *117*, 108391. [CrossRef]
18. Nguyen, H.; Huynh, H.; Tran, T.; Huynh, H. Explanation of the Convolutional Neural Network Classifying Chest X-Ray Images Supporting Pneumonia Diagnosis. *EAI Endorsed Trans. Context-Aware Syst. Appl.* **2020**, *7*, e3. [CrossRef]

19. Lo, S.-H.; Yin, Y. A Novel Interaction-Based Methodology towards Explainable AI with Better Understanding of Pneumonia Chest X-Ray Images. *Discov. Artif. Intell.* **2021**, *1*, 16. [\[CrossRef\]](#)
20. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
21. Philbrick, K.A.; Yoshida, K.; Inoue, D.; Akkus, Z.; Kline, T.L.; Weston, A.D.; Korfiatis, P.; Takahashi, N.; Erickson, B.J. What Does Deep Learning See? Insights From a Classifier Trained to Predict Contrast Enhancement Phase From CT Images. *AJR Am. J. Roentgenol.* **2018**, *211*, 1184–1193. [\[CrossRef\]](#)
22. Shen, Y.; Huang, X. A Comparative Visualization Analysis of Neural Network Models Using Grad-CAM. *Sci. Technol. Eng. Chem. Environ. Prot.* **2024**, *1*, 10. [\[CrossRef\]](#)
23. Xiao, M.; Zhang, L.; Shi, W.; Liu, J.; He, W.; Jiang, Z. A Visualization Method Based on the Grad-CAM for Medical Image Segmentation Model. In Proceedings of the 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, 23–26 September 2021; pp. 242–247.
24. Ali, A.A. Interpretable Deep Learning Framework for COVID-19 Detection: Grad-CAM Integration with Pre-Trained CNN Models on Chest X-Ray Images. *Int. J. Sci. Res. Sci. Eng. Technol.* **2025**, *12*, 153–163. [\[CrossRef\]](#)
25. Saporta, A.; Gui, X.; Agrawal, A.; Pareek, A.; Truong, S.Q.H.; Nguyen, C.D.T.; Ngo, V.-D.; Seekins, J.; Blankenberg, F.G.; Ng, A.Y.; et al. Benchmarking Saliency Methods for Chest X-Ray Interpretation. *Nat. Mach. Intell.* **2022**, *4*, 867–878. [\[CrossRef\]](#)
26. Suara, S.; Jha, A.; Sinha, P.; Sekh, A.A. Is Grad-CAM Explainable in Medical Images? In *Proceedings of the Computer Vision and Image Processing*; Kaur, H., Jakhethiya, V., Goyal, P., Khanna, P., Raman, B., Kumar, S., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 124–135.
27. Web of Science. Available online: <https://www.webofscience.com> (accessed on 18 April 2025).
28. Radočaj, P.; Radočaj, D.; Martinović, G. Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays. *Big Data Cogn. Comput.* **2025**, *9*, 25. [\[CrossRef\]](#)
29. Naydenova, E.; Tsanas, A.; Casals-Pascual, C.; De Vos, M. Smart Diagnostic Algorithms for Automated Detection of Childhood Pneumonia in Resource-Constrained Settings. In Proceedings of the 2015 IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA, 8–11 October 2015; pp. 377–384.
30. Luján-García, J.E.; Yáñez-Márquez, C.; Villuendas-Rey, Y.; Camacho-Nieto, O. A Transfer Learning Method for Pneumonia Classification and Visualization. *Appl. Sci.* **2020**, *10*, 2908. [\[CrossRef\]](#)
31. Panwar, H.; Gupta, P.K.; Siddiqui, M.K.; Morales-Menendez, R.; Bhardwaj, P.; Singh, V. A Deep Learning and Grad-CAM Based Color Visualization Approach for Fast Detection of COVID-19 Cases Using Chest X-Ray and CT-Scan Images. *Chaos Solitons Fractals* **2020**, *140*, 110190. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Zebin, T.; Rezvy, S. COVID-19 Detection and Disease Progression Visualization: Deep Learning on Chest X-Rays for Classification and Coarse Localization. *Appl. Intell. Dordr. Neth.* **2021**, *51*, 1010–1021. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Rahman, M.F.; Tseng, T.-L.; Pokojovy, M.; McCaffrey, P.; Walser, E.; Moen, S.; Vo, A.; Ho, J.C. Machine-Learning-Enabled Diagnostics with Improved Visualization of Disease Lesions in Chest X-Ray Images. *Diagnostics* **2024**, *14*, 1699. [\[CrossRef\]](#)
34. Mohagheghi, S.; Alizadeh, M.; Safavi, S.M.; Foruzan, A.H.; Chen, Y.-W. Integration of CNN, CBMIR, and Visualization Techniques for Diagnosis and Quantification of Covid-19 Disease. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1873–1880. [\[CrossRef\]](#)
35. Owais, M.; Yoon, H.S.; Mahmood, T.; Haider, A.; Sultan, H.; Park, K.R. Light-Weighted Ensemble Network with Multilevel Activation Visualization for Robust Diagnosis of COVID19 Pneumonia from Large-Scale Chest Radiographic Database. *Appl. Soft Comput.* **2021**, *108*, 107490. [\[CrossRef\]](#)
36. Misra, D. Mish: A Self Regularized Non-Monotonic Activation Function. *arXiv* **2019**, arXiv:1908.08681.
37. Radočaj, P.; Radočaj, D.; Martinović, G. Pediatric Pneumonia Recognition Using an Improved DenseNet201 Model with Multi-Scale Convolutions and Mish Activation Function. *Algorithms* **2025**, *18*, 98. [\[CrossRef\]](#)
38. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131.e9. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Keras 3 API Documentation. Available online: <https://keras.io/api/> (accessed on 24 September 2024).
40. Module: Tf | TensorFlow v2.16.1. Available online: https://www.tensorflow.org/api_docs/python/tf (accessed on 4 May 2024).
41. Liu, J.; Zhang, L.; Guo, A.; Gao, Y.; Zheng, Y. Multi-Scale Feature Fusion Convolutional Neural Network for Multi-Modal Medical Image Fusion. In Proceedings of the 2023 4th International Conference on Computing, Networks and Internet of Things, Xiamen, China, 26–28 May 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 913–917.
42. Yi, R.; Tang, L.; Tian, Y.; Liu, J.; Wu, Z. Identification and Classification of Pneumonia Disease Using a Deep Learning-Based Intelligent Computational Framework. *Neural Comput. Appl.* **2023**, *35*, 14473–14486. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Charan, K.S.; Krishna, O.V.; Sai, P.V.; Ilavarasi, A.K. Transfer Learning Based Multi-Class Lung Disease Prediction Using Textural Features Derived From Fusion Data. *IEEE Access* **2024**, *12*, 108248–108262. [\[CrossRef\]](#)

44. Bhatt, D.; Patel, C.; Talsania, H.; Patel, J.; Vaghela, R.; Pandya, S.; Modi, K.; Ghayvat, H. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. *Electronics* **2021**, *10*, 2470. [\[CrossRef\]](#)
45. Sanghvi, H.A.; Patel, R.H.; Agarwal, A.; Gupta, S.; Sawhney, V.; Pandya, A.S. A Deep Learning Approach for Classification of COVID and Pneumonia Using DenseNet-201. *Int. J. Imaging Syst. Technol.* **2023**, *33*, 18–38. [\[CrossRef\]](#)
46. Roy, P.; Efat, A.H.; Hasan, S.M.; Srizon, A.Y.; Hossain, M.R.; Faruk, M.F.; Al Mamun, M. Multi-Scale Feature Fusion Framework Based on Attention Integrated Customized DenseNet201 Architecture for Multi-Class Skin Lesion Detection. In Proceedings of the 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON), Rajshahi, Bangladesh, 12–13 September 2024; pp. 496–501.
47. Yuan, H.; Cheng, J.; Wu, Y.; Zeng, Z. Low-Res MobileNet: An Efficient Lightweight Network for Low-Resolution Image Classification in Resource-Constrained Scenarios. *Multimed. Tools Appl.* **2022**, *81*, 38513–38530. [\[CrossRef\]](#)
48. Yang, Y.; Mei, G.; Piccialli, F. A Deep Learning Approach Considering Image Background for Pneumonia Identification Using Explainable AI (XAI). *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2024**, *21*, 857–868. [\[CrossRef\]](#)
49. Desai, S.; Ramaswamy, H.G. Ablation-CAM: Visual Explanations for Deep Convolutional Network via Gradient-Free Localization. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA, 1–5 March 2020; pp. 983–991.
50. Namatëvs, I. Deep Convolutional Neural Networks: Structure, Feature Extraction and Training. *Inf. Technol. Manag. Sci.* **2017**, *20*, 40–47. [\[CrossRef\]](#)
51. Yan, T.; Wong, P.K.; Ren, H.; Wang, H.; Wang, J.; Li, Y. Automatic Distinction between COVID-19 and Common Pneumonia Using Multi-Scale Convolutional Neural Network on Chest CT Scans. *Chaos Solitons Fractals* **2020**, *140*, 110153. [\[CrossRef\]](#)
52. Sarkar, O.; Islam, M.R.; Syfullah, M.K.; Islam, M.T.; Ahamed, M.F.; Ahsan, M.; Haider, J. Multi-Scale CNN: An Explainable AI-Integrated Unique Deep Learning Framework for Lung-Affected Disease Classification. *Technologies* **2023**, *11*, 134. [\[CrossRef\]](#)
53. Yang, C.; Wang, Y.; Wang, X.; Geng, L. A Stride-Based Convolution Decomposition Method to Stretch CNN Acceleration Algorithms for Efficient and Flexible Hardware Implementation. *IEEE Trans. Circuits Syst. Regul. Pap.* **2020**, *67*, 3007–3020. [\[CrossRef\]](#)
54. Younesi, A.; Ansari, M.; Fazli, M.; Ejlali, A.; Shafique, M.; Henkel, J. A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends. *IEEE Access* **2024**, *12*, 41180–41218. [\[CrossRef\]](#)
55. Ayachi, R.; Afif, M.; Said, Y.; Atri, M. Strided Convolution Instead of Max Pooling for Memory Efficiency of Convolutional Neural Networks. In *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18)*; Bouhlel, M.S., Rovetta, S., Eds.; Springer International Publishing: Cham, Switzerland, 2020; Volume 1; pp. 234–243.
56. Zaniolo, L.; Marques, O. On the Use of Variable Stride in Convolutional Neural Networks. *Multimed. Tools Appl.* **2020**, *79*, 13581–13598. [\[CrossRef\]](#)
57. Fränti, P.; Marinescu-Istodor, R. Soft Precision and Recall. *Pattern Recognit. Lett.* **2023**, *167*, 115–121. [\[CrossRef\]](#)
58. Khan, S.I.; Shahriar, A.; Karim, R.; Hasan, M.; Rahman, A. MultiNet: A Deep Neural Network Approach for Detecting Breast Cancer through Multi-Scale Feature Fusion. *J. King Saud Univ.—Comput. Inf. Sci.* **2022**, *34*, 6217–6228. [\[CrossRef\]](#)
59. Franquet, T. Imaging of Pulmonary Viral Pneumonia. *Radiology* **2011**, *260*, 18–39. [\[CrossRef\]](#)
60. Garg, M.; Prabhakar, N.; Kiruthika, P.; Agarwal, R.; Aggarwal, A.; Gulati, A.; Khandelwal, N. Imaging of Pneumonia: An Overview. *Curr. Radiol. Rep.* **2017**, *5*, 16. [\[CrossRef\]](#)
61. Franquet, T. Imaging of Community-Acquired Pneumonia. *J. Thorac. Imaging* **2018**, *33*, 282. [\[CrossRef\]](#)
62. Nillmani; Jain, P.K.; Sharma, N.; Kalra, M.K.; Viskovic, K.; Saba, L.; Suri, J.S. Four Types of Multiclass Frameworks for Pneumonia Classification and Its Validation in X-Ray Scans Using Seven Types of Deep Learning Artificial Intelligence Models. *Diagnostics* **2022**, *12*, 652. [\[CrossRef\]](#)
63. Li, Y.; Zhang, Z.; Dai, C.; Dong, Q.; Badrigilan, S. Accuracy of Deep Learning for Automated Detection of Pneumonia Using Chest X-Ray Images: A Systematic Review and Meta-Analysis. *Diagnostics* **2020**, *123*, 103898. [\[CrossRef\]](#)
64. Chapman, W.W.; Fizman, M.; Chapman, B.E.; Haug, P.J. A Comparison of Classification Algorithms to Automatically Identify Chest X-Ray Reports That Support Pneumonia. *J. Biomed. Inform.* **2001**, *34*, 4–14. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Wu, Y.; Rocha, B.M.; Kaimakamis, E.; Cheimariotis, G.-A.; Petmezas, G.; Chatzis, E.; Kilintzis, V.; Stefanopoulos, L.; Pessoa, D.; Marques, A.; et al. A Deep Learning Method for Predicting the COVID-19 ICU Patient Outcome Fusing X-Rays, Respiratory Sounds, and ICU Parameters. *Expert Syst. Appl.* **2024**, *235*, 121089. [\[CrossRef\]](#)
66. Wehbe, R.M.; Sheng, J.; Dutta, S.; Chai, S.; Dravid, A.; Barutcu, S.; Wu, Y.; Cantrell, D.R.; Xiao, N.; Allen, B.D.; et al. DeepCOVID-XR: An Artificial Intelligence Algorithm to Detect COVID-19 on Chest Radiographs Trained and Tested on a Large U.S. Clinical Data Set. *Radiology* **2021**, *299*, E167–E176. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

CURRICULUM VITAE

Petra Radočaj was born on June 20, 1998, in Našice, Croatia. She had a direct enrollment to the Faculty of Electrical Engineering, Computer Science and Information Technologies Osijek (FERIT Osijek) in the academic year 2017/2018. She earned a Bachelor's degree in Computer Science, graduating in 2020. Her final thesis focused on recognizing plant leaf diseases using machine learning based on images. She continued her studies at FERIT Osijek, completing a Master's degree in Computer Science – Software Engineering in 2022. Her master's thesis explored the recognition of tomato diseases using transfer deep machine learning with a developed module. Following her master's studies, she pursued a doctoral degree in Computer Science at FERIT Osijek, enrolling in 2022. During her studies, she was a student teaching assistant, assisting with several courses.

She has been a scholarship recipient every year since her third year of high school, continuing through her doctoral studies. In recognition of her academic achievements, she received the Faculty Award for excellent success during her master's studies in Software Engineering from FERIT Osijek in 2022. In 2021, she was awarded the Rector's Award by Josip Juraj Strossmayer University of Osijek for her seminar paper on "Detection of pancreatic cancer using machine learning methods with urine biomarkers". She also received recognition for successful academic performance in her postgraduate university studies in 2025.

Her professional experience includes her current role as a Business Analyst/Product Owner at Layer d.o.o. in Osijek. In this role, she contributes to innovative, tailor-made client projects, focusing on understanding business processes, identifying challenges and opportunities, informing key decisions, and applying machine learning. She also plays a crucial role in guiding project development from its initial phase to successful delivery. Previously, she worked as a software developer and architect, including on a project for personalized medicine in oncology as part of a national cancer early detection program. She also worked on maintaining and improving embedded systems and conducting research on data analysis for cancer patient survival prediction using deep machine learning. She also holds a qualification in "Acquisition of Micro-qualifications for the Preparation and Implementation of Projects Financed by EU Funds" from the University Algebra Bernays, which she completed in 2024. She has obtained several certifications in areas such as accelerated computing, embedded systems software and hardware, neural networks and deep learning.

She is proficient in English at the B2 level and actively uses it. She also has basic knowledge of Italian and German.

She is the author and co-author of scientific and professional publications as follows:

Papers indexed in Web of Science – Current Contents (CC) and Science Citation Index

Expanded (SCIE) Q1/Q2:

1. **Radočaj, P.**, Martinović, G., Emotion Recognition in Autistic Children Through Facial Expressions Using Advanced Deep Learning Architectures. *Applied Sciences*, 15 (2025), 17., 9555.

2. **Radočaj P.**, Martinović G., Interpretable Deep Learning for Pediatric Pneumonia Diagnosis Through Multi-Phase Feature Learning and Activation Patterns. *Electronics*, 14 (2025), 9., 1899.
3. **Radočaj P.**, Radočaj, D., & Martinović, G., Pediatric Pneumonia Recognition Using an Improved DenseNet201 Model with Multi-Scale Convolutions and Mish Activation Function. *Algorithms*, 18 (2025), 2., 98.
4. **Radočaj P.**, Radočaj, D., & Martinović, G., Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-rays. *Big Data and Cognitive Computing*, 9 (2025), 2., 25.
5. **Radočaj P.**, Radočaj, D., & Martinović, G., Image-based leaf disease recognition using transfer deep learning with a novel versatile optimization module. *Big Data and Cognitive Computing*, 8 (2024), 6., 52.
6. Radočaj, D., Gašparović, M., **Radočaj P.**, & Jurišić, M., Geospatial prediction of total soil carbon in European agricultural land based on deep learning. *Science of the Total Environment*, 912 (2024), 169647.