Exploring Deep Learning Architectures for Histopathologic Cancer Detection: A Comparative Analysis

Alperen Fatih Sakarya & Burak Polat

Hasso-Plattner-Institute & University of Potsdam

August 2023

Abstract

Today, deep learning and machine learning are being used in numerous industries. One of these industries in particular is the medical industry. With the recent technological developments, we started to use computers more and more to help us tackle with the challenges we come across. Recent researches have effectively applied Convolutional Neural Networks to identify lung cancer in histopathological lung images. This study focuses on contrasting two network designs for the detection of metastatic tissue in histopathological lymph node scans. Baseline CNN and AlexNet. The PatchCamelyon dataset was employed for performance evaluation. The Baseline CNN demonstrated the highest accuracy (91%) and recall (86.5%) in our experiment.

Contents

1 Introduction

Imagine a world where artificial intelligence-based algorithms are used to understand, recognise and extract important information from the images in our lives. This is where "Convolutional Neural Networks" or CNNs come into play. CNNs work with deep learning principles as algorithms that perform complex tasks on photographs. They play an important role in cancer diagnosis and treatment, especially in the field of medicine. Because cancer stands out as a major health problem in human life. Cancer is a disease that can develop in different parts of the body and is caused by the uncontrolled proliferation of cells. There are different types of cancers and each may require different treatment and diagnostic methods. At this point, artificial intelligence-based Convolutional Neural Networks (CNNs) have an important role in cancer diagnosis [5]. Histopathological cancer detection is a method in which cancer cells are detected by examining them under a microscope. This is a long and laborious process in which expert pathologists carefully examine tissue samples to identify cancer cells. This is where the power of CNNs comes into play. CNNs are a special type of deep learning models and have achieved great success, especially in image processing [3]. For example, a CNN architecture such as AlexNet was a huge success in the ImageNet competition in 2012 and revolutionised the field of image recognition. AlexNet is characterised by its deep architecture, ReLU activation functions, and the capabilities gained by training on large data sets. When used for cancer diagnosis, AlexNet and other CNN architectures can automatically detect important features in histopathological images and accelerate cancer detection [1]. This is vital for early detection and more effective treatment planning. In addition to the benefits that CNNs bring to cancer diagnosis in the medical field, it is possible to see this technology in our daily lives. Social media platforms can use CNNs for automatic tagging and content analysis. Security cameras and security systems in traffic can use CNNs to detect and intervene in potential dangers. In summary, CNNs provide great benefits in visual analysis and diagnosis, both in important tasks such as cancer diagnosis in the medical field and in our daily lives. Especially in a vital field such as cancer detection, the use of deep learning techniques and CNNs can speed up the diagnostic processes and increase the life chances of patients. At the same time, we can create a smarter and safer environment in our daily lives thanks to this technology.

2 Related Work

In the realm of cancer detection and medical image analysis, Convolutional Neural Networks (CNNs) have gained significant attention, proving to be powerful tools for various tasks. One of the pioneering CNN architectures in this domain is AlexNet, which revolutionized the field with its deep architecture and use of ReLU activation functions. Several studies have explored the application of AlexNet and CNNs for histopathological image analysis, particularly in cancer detection. For instance, in a recent study, researchers utilized AlexNet to

classify histopathological lung images into different lung cancer types, including benign tissue, adenocarcinoma, and squamous cell carcinoma [13]. The CNN achieved remarkable accuracy rates during training and validation, highlighting its potential as a reliable tool for lung cancer classification [6]. Furthermore, the versatility of CNNs is evident in studies focused on osteosarcoma tumor classification. A novel CNN architecture was designed, comprising eight layers, including stacked convolutional layers and max pooling layers for feature extraction. By employing data augmentation techniques, the CNN demonstrated a high average classification accuracy for differentiating between viable tumor and necrosis in histopathologic osteosarcoma images [9]. In addition to lung cancer and osteosarcoma, CNNs have shown promise in diagnosing breast cancer and colon adenocarcinomas. One study aimed to develop a computer-aided diagnosis method for these malignancies by analyzing digital pathology images. A shallow neural network design was used to categorize histological slides of lung and colon tumors into their respective classes. The proposed method achieved impressive accuracy rates, indicating the potential for accurate cancer diagnosis using CNN-based approaches [8]. Transfer learning, a technique that involves modifying pre-trained CNN models for specific tasks, has also been extensively explored. A study focused on histopathologic breast cancer identification used transfer learning and deep feature extraction techniques with AlexNet and Vgg16 models [2]. The CNN models were then fine-tuned using Support Vector Machines (SVM) for classification. The study demonstrated that transfer learning outperformed deep feature extraction and SVM classification, showcasing its superiority in breast cancer diagnosis [10]. Moreover, CNN architectures trained on a combination of facial and natural images have been applied to biomedical image analysis, particularly in cell nucleus classification. Empirical findings revealed that transfer learning from pre-trained networks significantly improved classification performance, while reducing training time compared to starting from scratch. Deep CNN architectures, such as VGG-16, GenderNet, GoogleNet, and AlexNet, have been employed to achieve accurate classification of cell nuclei in histopathological images [4]. In conclusion, the related work demonstrates the immense potential of CNNs, including AlexNet, in cancer detection and medical image analysis. These deep learning models have proven to be versatile and effective tools for classifying various cancer types and offer valuable insights to pathologists and medical professionals, contributing to improved diagnostic accuracy and patient outcomes.

3 Theoretical Background

3.1 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are a specialized class of artificial neural networks designed to process and analyze grid-like data, with a primary focus on image analysis. These powerful networks have demonstrated remarkable success in various computer vision tasks, including image classification, object detection, and medical image analysis, such as cancer detection. At the heart of a CNN are the convolutional layers, which use learnable filters (kernels) to scan the input image. These filters extract specific features by sliding over the image and performing element-wise multiplication and summation operations. The resulting feature maps highlight different patterns, edges, and textures present in the input image. The ability to automatically learn these features makes CNNs highly effective in recognizing complex visual patterns and objects. Following the convolutional layers, pooling layers are employed to down-sample the feature maps. Pooling operations, such as max-pooling, reduce the spatial dimensions of the data while preserving the most important information. This process not only reduces computational complexity but also enhances the network's ability to generalize by extracting the most salient features. To facilitate decision-making, fully connected layers are employed at the end of the CNN architecture. These layers take the flattened output from the previous layers and apply a series of weighted summations, leading to final predictions or classifications. The learned weights and biases during training enable the network to accurately categorize the input data into different classes. One of the remarkable aspects of CNNs is their hierarchical feature learning. The initial layers detect basic features like lines and curves, while deeper layers progressively recognize more intricate and abstract features, allowing for a rich representation of the input data. CNNs have found widespread application in numerous fields, including healthcare. In cancer detection, CNNs analyze medical images, such as histopathological slides, to identify potential cancerous regions with high accuracy. Their ability to discern subtle patterns and anomalies in medical images can aid in early diagnosis and treatment planning, leading to improved patient outcomes. In recent years, advancements in CNN architectures and the availability of large-scale datasets have further boosted their performance. Transfer learning, a technique that leverages pre-trained CNN models on large image datasets, enables the application of CNNs to medical image analysis with limited data. In conclusion, Convolutional Neural Networks have revolutionized the field of computer vision and significantly impacted medical image analysis. Their capacity to learn intricate features and patterns has propelled advancements in cancer detection and other critical healthcare applications, paving the way for more precise diagnoses and enhanced medical decision-making. [3]

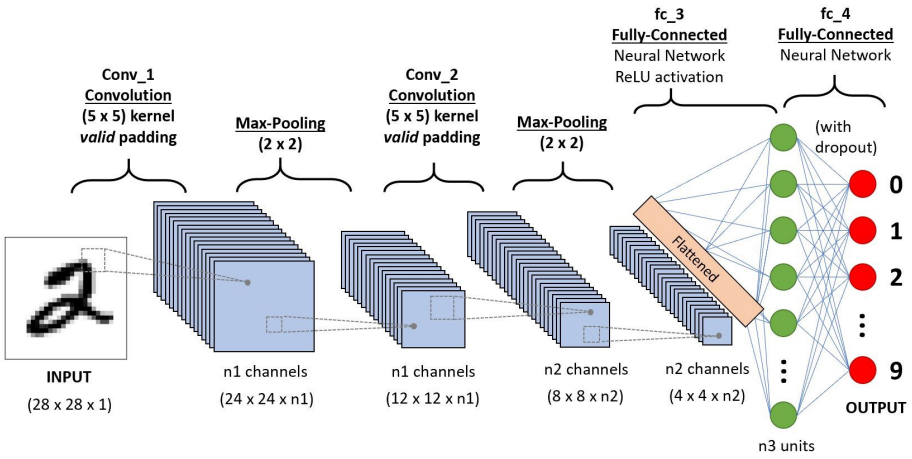

Figure 1: An exemplar architecture of a CNN model [11]

3.2 Transfer Learning and Alexnet

Transfer learning is a powerful technique in deep learning that allows us to leverage knowledge gained from one task or domain and apply it to a new, related task. It is particularly beneficial when data for the target task is limited, as it reduces the need for a large amount of labeled data. The process of transfer learning involves using a pre-trained CNN as a starting point for the new task. We first remove the fully connected layers responsible for the original task and keep the convolutional layers as feature extractors. This step is known as "Feature Extraction," where we extract high-level feature representations from the input data. Once we have obtained the feature representations, we add new fully connected layers specifically designed for the target task. These new layers are initialized randomly, and the model is fine-tuned by updating their weights while keeping the convolutional layers frozen or with a small learning rate. This step is referred to as "Fine-Tuning," allowing the model to adapt its learned features to the nuances of the new task.Transfer learning offers several advantages, including faster training times, better generalization, and improved performance on the target task. In medical image analysis, such as cancer detection, transfer learning has proven to be especially valuable due to the scarcity of labeled medical data. When it comes to models built on the foundation of transfer learning and further advancements in deep learning, AlexNet has become a prominent reference point. While AlexNet played a crucial role in popularizing deep learning for image analysis, subsequent researchers and developers have built upon its principles to create even more sophisticated and efficient models. These newer models have been tailored to tackle specific tasks and domains, showcasing the continuous evolution of transfer learning in the field of medical image analysis, including cancer detection. In this context, the

AlexNet architecture stands out as a seminal CNN model that played a pivotal role in popularizing deep learning for image analysis. AlexNet, developed by Alex Krizhevsky, achieved a breakthrough by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. [7]

AlexNet's architecture comprises eight layers, including five convolutional layers followed by three fully connected layers. The incorporation of non-linear activation functions (ReLU) and dropout layers helped mitigate overfitting. With approximately 60 million parameters, AlexNet was one of the first large-scale CNNs, showcasing the power of deep learning in handling complex visual data. The key strengths of AlexNet lie in its ability to learn hierarchical features, capturing both low-level patterns like edges and high-level features like object parts and textures. Its depth and width enable it to model intricate relationships in the data, making it highly effective in image recognition tasks. However, one of the main drawbacks of AlexNet is its computational complexity. The large number of parameters requires considerable computation resources, making training and deploying AlexNet more time and resource-intensive compared to more lightweight architectures. In conclusion, the combination of transfer learning and AlexNet has significantly advanced the field of deep learning and medical image analysis, particularly in cancer detection. By leveraging transfer learning techniques and building on the strengths of AlexNet, researchers and clinicians can achieve accurate and efficient diagnosis, ultimately leading to improved patient outcomes. [15]

3.3 Max-Pooling & Average-Pooling

Max pooling and average pooling are both pooling operations commonly used in convolutional neural networks (CNNs) as part of the feature extraction process. They help reduce the spatial dimensions of the input data while retaining important features. As the names suggest, Max-Pooling selects the highest value within a window, capturing key features and promoting translation invariance while Average-Pooling calculates the average value within a window, capturing general patterns and reducing sensitivity to outliers. Both techniques aid in managing computational complexity, promoting spatial hierarchy, and preventing overfitting in CNNs. The choice between them depends on the task and data characteristics. [10]

3.4 Dropout

Dropout is a technique in deep learning where during training, random neurons are temporarily turned off, preventing the network from relying too heavily on specific ones. This prevents overfitting, helps the network generalize better, and acts like training multiple models at once. It's an efficient way to improve model performance and reduce training time. [14]

3.5 Batch Normalization

Batch Normalization is a technique in deep learning that normalizes the inputs of a layer within a mini-batch during training. It helps stabilize and speed up training by reducing internal covariate shifts, allowing the network to learn more quickly and generalize better. It also acts as a form of regularization, enhancing model performance and convergence. [12]

4 Architecture

In this chapter, we will delve into the details of how the Histopathologic Cancer Detection models were trained. The project utilized two distinct approaches: the Baseline Convolutional Neural Network (CNN) and AlexNet Transfer Learning, which were explained in the previous chapter.

In the models, we used different kinds of activation functions and adjusted the models depths according to the results we've gotten from experimenting with different architectures. Here's a complete rundown of the models:

4.1 Baseline CNN Architecture

Our Baseline CNN model consists of multiple layers, including convolutional layers, activation functions, pooling layers, batch normalization, and fully connected layers.

1. **Input Layer**: The input layer takes images with a dimension of img_dim (e.g., $3 \times 128 \times 128$). Each image has three channels (RGB).

2. **Convolutional Layer 1**: This layer performs the first convolution on the input image. It has 16 output channels and uses a 3×3 kernel. The output shape after convolution is $16 \times 126 \times 126$.

3. **ReLU Activation**: The Rectified Linear Unit (ReLU) activation function introduces non-linearity into the model, applied element-wise to the output of the previous layer.

4. **Max Pooling Layer 1**: This layer performs max pooling with a 2×2 kernel and a stride of 2. It reduces the spatial dimensions by half, resulting in an output shape of $16 \times 63 \times 63$.

5. **Batch Normalization 1**: Batch normalization helps stabilize and accelerate the training process by normalizing the input to each neuron.

6. **Convolutional Layer 2**: The second convolutional layer has 32 output channels and uses a 3×3 kernel. The output shape after convolution is $32 \times 61 \times 61$.

7. **Leaky ReLU Activation**: The Leaky ReLU activation function is an alternative to ReLU, allowing a small negative slope for negative inputs.

8. **Max Pooling Layer 2**: Similar to Max Pooling Layer 1, this layer performs max pooling with a 2×2 kernel and a stride of 2, resulting in an output shape of $32 \times 30 \times 30$.

9. **Batch Normalization 2**: Batch normalization is applied again to the output of the second convolutional layer.

10. **Convolutional Layer 3**: The third convolutional layer has 64 output channels and uses a 3×3 kernel. The output shape after convolution is $64 \times 28 \times 28$.

11. **Tanh Activation**: The hyperbolic tangent (Tanh) activation function introduces non-linearity and squashes the output between -1 and 1.

12. **Max Pooling Layer 3**: Max pooling is applied with a 2×2 kernel and a stride of 2, resulting in an output shape of $64 \times 14 \times 14$.

13. **Batch Normalization 3**: Batch normalization is applied to the output of the third convolutional layer.

14. **Global Average Pooling Layer**: This layer performs global average pooling, reducing the spatial dimensions to $64 \times 1 \times 1$.

15. **Batch Normalization 4**: Batch normalization is applied again to the output of the global average pooling layer.

16. **Flatten Layer**: The flatten layer reshapes the output to a 1D vector with 64 elements.

17. **Fully Connected Layer 1**: This fully connected layer has 32 neurons and applies the ReLU activation function.

18. **Dropout Layer 1**: Dropout is used to prevent overfitting by randomly setting a fraction of neurons' outputs to zero during training.

19. **Fully Connected Layer 2**: This fully connected layer has 16 neurons and applies the ReLU activation function.

20. **Dropout Layer 2**: Dropout is applied again to the output of the second fully connected layer.

21. **Fully Connected Layer 3**: This fully connected layer has 8 neurons and applies the ReLU activation function.

22. **Fully Connected Layer 4 (Output Layer)**: The output layer consists of a single neuron and uses the sigmoid activation function, which squashes the output between 0 and 1, representing the probability of the image being cancerous.

4.2 AlexNet Transfer Learning Architecture

The AlexNet Transfer Learning model is based on the well-known AlexNet architecture, which was pretrained on a large-scale image classification task (e.g., ImageNet). We adapt the AlexNet classifiers to suit the histopathologic cancer detection problem.

1. **Input Layer**: The input layer takes images with a dimension of *img_dim* (e.g., $3 \times 128 \times 128$). Each image has three channels (RGB).

2. **Pretrained AlexNet**: The model is instantiated with pretrained weights from AlexNet, obtained from a well-known source of weights (e.g., PyTorch torchvision). The weights are frozen to prevent updates during the training of the new layers.

3. **Adapting the AlexNet Classifiers**: - The second classifier of the original AlexNet, which had 4096 input features, is replaced with a new fully connected layer with 1024 neurons. This is done by modifying the first element of the *classifier* list of layers in AlexNet.

   - The fourth classifier, which had 4096 input features, is replaced with another fully connected layer with 256 neurons.

   - A Batch Normalization layer is added after the 256-neuron layer to normalize the input.

   - A Dropout layer is introduced after Batch Normalization to prevent overfitting.

   - A ReLU activation function is applied after Dropout to introduce nonlinearity.

4. **Output Layer**: - The final classifier of the pretrained AlexNet, which had 1000 output neurons for ImageNet's classification, is modified to have a single neuron for the binary classification task of histopathologic cancer detection.

   - Additional classifiers are introduced to further improve the model's performance: - A Dropout layer is added after the last classifier to prevent overfitting.

   - A ReLU activation function is applied after Dropout to introduce nonlinearity.

   - A fully connected layer with 64 neurons is added.

   - A ReLU activation function is applied after the 64-neuron layer.

   - A fully connected layer with 32 neurons is added.

   - A ReLU activation function is applied after the 32-neuron layer.

   - The final fully connected layer has a single neuron and uses the sigmoid activation function to squash the output between 0 and 1, representing the probability of the image being cancerous.

5 Training

In order to maintain consistent and comparable results, all architectures were trained using identical hyperparameters. The batch size was fixed at 64, and the initial learning rate was set to 0.001. The integration of a Learning Rate Scheduler was subsequently abandoned from the methodology due to concerns regarding potential interference with the experimental outcomes, notwithstanding its potential to enhance the results. An early stopping algorithm might still be implemented because the training accuracy stagnates after few epoches. Given the binary classification nature of the task, the preferred choice for the loss function was binary cross-entropy. Overall, each model underwent training for 5 epochs, amounting to a total of 25000 batches during the training process.

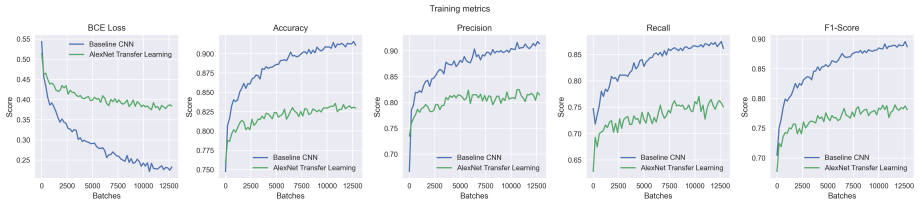

Figure 2: Comparison between models metrics while training

6 Evaluation

To assess the distinction between the two models, evaluation metrics were computed on the training and test dataset for each model. Specifically, accuracy, precision, recall, and F1-Score were used. Considering the medical nature of the classification task, recall can be considered as the most preferred metric among all of them, since false positives has less of an impact than false negative when regarding peoples health.

Model	**Loss**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Baseline CNN	0,2327	0,9110	0,9138	0,8615	0,8869
AlexNet Transfer Learning	0,3841	0,8296	0,8146	0,7502	0,7811

Table 1: Result of the models performance based on the last epoch of the training

Model	Accuracy	Precision	Recall	F1-Score
Baseline CNN	0,9077	0,9030	0,8651	0,8837
AlexNet Transfer Learning	0,8289	0,8008	0,7690	0,7846

Table 2: Result of the models performance based on testing

7 Results

The results of the experimental analysis showcase notable differences in the performance of the two employed models. The Baseline CNN exhibited a substantially higher recall rate, achieving an noticable 87% accuracy in correctly identifying cancerous instances within the histopathologic images. In contrast, the AlexNet Transfer Learning model demonstrated a slightly lower recall rate, achieving a commendable yet comparatively lower 77% accuracy in accurately detecting cancerous tissue. Overall accuracy of the models also differ noticeably with %91 and %83. This discrepancy highlights the efficacy of the Baseline CNN in identifying critical cases of malignancy, underscoring its potential superiority over the AlexNet Transfer Learning model in this specific histopathologic cancer detection context.

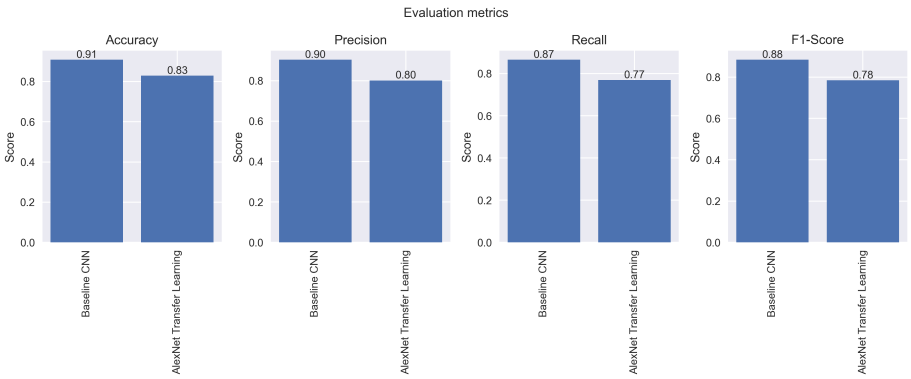

Figure 3: Comparison of the results of the metrics of each model

8 Further Research

Although we experimented continuously with the hyperparameters and with the depth of the models, there still seems to be a room for improvement. Especially further implementations to the model such as an Early Stopping Algorithm or a Learning Rate Scheduler might improve both the loss rate and the resource control. Furthermore changing the architecture of the Baseline CNN model and adding more layers or experimenting with the combination of different activation functions such as Swish may give different results.

9 Conclusion

A significant achievement has been reached in the field of medical pathology through the utilization of digitalized histopathology images to identify breast cancer. This advancement has not only opened up new opportunities for research, but it has also allowed the application of machine learning and deep learning techniques to explore previously uncharted domains. The objective of our study was to develop a method for detecting cancer in histopathological samples using various architectural designs. To be exact, deep learning models: Baseline CNN and AlexNet Transfer Learning. We started by increasing the depth of the preexisting models, implementing different combinations of activation functions and adding more batch normalization. Thus we achieved better results at detecting true positive cases, especially with AlexNet Transfer Learning. In future works, we can experiment with different architectures and improve our models.

References and citations

[1] Leila Abdelrahman et al. "Convolutional neural networks for breast cancer detection in mammography: A survey". In: *Computers in biology and medicine* 131 (2021), p. 104248.

[2] Dheeb Albashish et al. "Deep CNN Model based on VGG16 for Breast Cancer Classification". In: *2021 International Conference on Information Technology (ICIT)*. 2021, pp. 805–810. DOI: 10.1109/ICIT52682.2021.9491631.

[3] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology (ICET)*. 2017, pp. 1–6. DOI: 10.1109/ICEngTechnol.2017.8308186.

[4] Neslihan Bayramoglu and Janne Heikkilä. "Transfer Learning for Cell Nuclei Classification in Histopathology Images". In: *Computer Vision – ECCV 2016 Workshops*. Ed. by Gang Hua and Hervé Jégou. Cham: Springer International Publishing, 2016, pp. 532–539. ISBN: 978-3-319-49409-8.

[5] Sumaiya Dabeer, Maha Mohammed Khan, and Saiful Islam. "Cancer diagnosis in histopathological image: CNN based approach". In: *Informatics in Medicine Unlocked* 16 (2019), p. 100231. ISSN: 2352-9148. URL: https://www.sciencedirect.com/science/article/pii/S2352914819301133.

[6] Bijaya Kumar Hatuwal and Himal Chand Thapa. "Lung cancer detection using convolutional neural network on histopathological images". In: *Int. J. Comput. Trends Technol* 68.10 (2020), pp. 21–24.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[8] Sanidhya Mangal, Aanchal Chaurasia, and Ayush Khajanchi. "Convolution neural networks for diagnosing colon and lung cancer histopathological images". In: *arXiv preprint arXiv:2009.03878* (2020).

[9] Rashika Mishra et al. "Convolutional neural network for histopathological analysis of osteosarcoma". In: *Journal of Computational Biology* 25.3 (2018), pp. 313–325.

[10] Rajendran Nirthika et al. "Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study". In: *Neural Computing and Applications* 34.7 (2022), pp. 5321–5347.

[11] Phani Ratan. *What is the Convolutional Neural Network Architecture?* Oct. 2020. URL: https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/.

[12] Shibani Santurkar et al. "How does batch normalization help optimization?" In: *Advances in neural information processing systems* 31 (2018).

[13] Prabira Kumar Sethy et al. "Lung cancer histopathological image classification using wavelets and AlexNet". In: *Journal of X-Ray Science and Technology* Preprint (2023), pp. 1–11.

[14] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[15] Wikipedia contributors. *AlexNet — Wikipedia, The Free Encyclopedia*. [Online; accessed 9-August-2023]. 2023. URL: https://en.wikipedia.org/w/index.php?title=AlexNet&oldid=1166474161.