Prediction of Cancer Metastasis in Lymph Node Whole-slide Image Patches with CNN

[redacted] Final Project
Team member: Shengyu Gao
Email: [redacted]
ID: [redacted]

Abstract

Breast cancer diagnosis can be significantly improved with the aid of computational methods that increase the accuracy and efficiency of detecting metastatic spread in sentinel lymph node biopsies. This study reports on the development of a deep learning model trained on the PatchCamelyon (PCam) dataset to classify histopathologic images of sentinel lymph node sections into metastatic and non-metastatic tissues. Utilizing a Convolutional Neural Network (CNN) architecture with sequential layers for feature extraction and classification, we applied batch normalization, max-pooling, and dropout strategies to enhance learning and prevent overfitting. The model was trained and validated across multiple epochs, achieving an accuracy of 82.79% on the test set. The confusion matrix revealed a higher rate of false negatives, indicating areas for potential model improvement. Despite these challenges, the model's performance suggests that it could serve as a valuable tool to assist pathologists in the timely and accurate diagnosis of breast cancer, potentially reducing the workload and improving patient outcomes. The study underscores the promise of deep learning in medical image analysis and the need for ongoing research to refine these models for clinical application.

Introduction

Breast cancer, one of the most common cancers worldwide, presents significant challenges in medical diagnostics [1]. Early detection and accurate diagnosis are crucial for effective treatment and patient care. The presence of lymph node metastases is one of the most important factors in breast cancer prognosis. The most common way to assess regional lymph node status is the sentinel lymph node procedure. The sentinel lymph node is the most likely lymph node to contain metastasized cancer cells and is excised, histopathologically processed, and examined by a pathologist. This tedious examination

process is time-consuming and can lead to small metastases being missed. However, recent advances in whole-slide imaging and machine learning have opened an avenue for analysis of digitized lymph node sections with computer algorithms.

State-of-the-art techniques in medical image analysis include models like U-Net and various adaptations of CNN architectures, demonstrating high accuracy in tasks such as tumor detection, segmentation, and classification [2]. These models leverage the power of deep learning to capture complex features in medical images, which are often not discernible by the human eye. The success of these models hinges on their ability to learn hierarchical feature representations, making them exceptionally well-suited for analyzing visual data such as pathology images. Recent advancements also include the incorporation of transfer learning, where networks pre-trained on large datasets (like ImageNet) are fine-tuned for specific medical imaging tasks, leading to improved performance even with relatively smaller datasets [3].

To investigate the potential of machine learning algorithms for detection of metastases in SLN slides and compare these with the diagnoses of pathologists, the Cancer Metastases in Lymph Nodes Challenge 2016 (CAMELYON16) competition was organized [4]. The resulted PatchCamelyon (PCam) dataset is a comprehensive collection of histopathologic scan images of whole-slide images of resected sentinel axillary lymph nodes (SLNs) from breast cancer patients and has become a benchmark in this realm. It consists of small patches extracted from larger digital pathology scans, where each patch is labeled for the presence of metastatic tissue. This dataset provides an excellent opportunity for developing and testing advanced image analysis algorithms. The primary objective of this project is to develop a CNN-based model to classify patches in the PCam dataset as either normal or indicative of metastatic cancer. By leveraging the capabilities of CNNs, this project aims to achieve high accuracy in identifying metastatic tissue, thereby contributing to the broader goal of enhancing diagnostic procedures in digital pathology.

Methodology

Dataset

The PCam dataset [5] comprises 327,680 color images (96x96 pixels each) of lymph node sections, each annotated with a binary label indicating the presence of metastatic tissue. This large-scale dataset is obtained by processing whole-slide images of sentinel lymph node sections, which are then partitioned into non-overlapping patches. Each patch is labeled by the presence (1) or absence (0) of metastatic tissue in the central 32x32 pixel region, ensuring a clear focus for classification tasks. The images in PCam are

preprocessed to a uniform size suitable for direct input into convolutional neural network models. The dataset has already been split into predefined training, validation, and test sets. We employed the H5Dataset class to handle the Pcam dataset.

Image Handling and Transformations

To augment the training data and improve model robustness, we applied random horizontal and vertical flips transformation to increase the diversity of the training data, simulating various orientations of histopathologic images. We also normalized the image pixel values to have a mean of 0.5 and a standard deviation of 0.5. This standardization is crucial for optimizing the training process, ensuring consistent scale across different images. For the test dataset, we apply similar normalization but exclude random flipping to maintain the original orientation of the test images.

Model Architecture

Overview

The CNN architecture is designed with the following components: four convolutional layers with increasing output channels (16, 32, 64, 128) capture a hierarchy of features from the input images; each convolutional layer is followed by batch normalization to stabilize learning and improve convergence; max-pooling is applied after each convolutional layer, which reduces the spatial dimensions by half, increasing the field of view and reducing the computational load; to prevent overfitting, dropout layers randomly zero out a fraction of the outputs from neurons; the network concludes with four fully connected layers, progressively focusing the model's learning on distinguishing features relevant to cancer detection. The network output is transformed through sigmoid function to obtain a binary classification indicating the presence or absence of metastatic tissue.

Convolutional Layers

The network comprises four convolutional layers, each serving as a feature detector. The first layer has 16 output channels, and each subsequent layer doubles the number of channels: 32, 64, and finally 128. Mathematically, the convolutional operation in the i^{th} layer can be expressed as: $F_i(x) = \text{ReLU}(W_i * x + b_i)$, where $*$ denotes the convolution operation, x is the input, W_i are the weights, b_i is the bias, and ReLU is the rectified linear activation function. The choice of increasing channels is to allow the network to learn increasingly complex features with depth.

Batch Normalization

Following each convolutional layer, batch normalization is applied, which normalizes the

output to have a mean of zero and a standard deviation of one. It can be formalized as: $BN_i(x) = \gamma \left(\frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \right) + \beta$, with μ_B and σ_B^2 representing the mini-batch mean and variance, respectively, γ and β being learnable parameters of scale and shift, and ϵ a constant added for numerical stability. This normalization aids in mitigating the internal covariate shift, thus expediting the training phase.

Max-Pooling

Each max-pooling layer reduces the spatial size of the representation, thus diminishing the number of parameters and computation in the network. The operation can be denoted as: $P_i(x) = \max_{(a,b)\in N} x(a, b)$, where N is the neighborhood defined by the pooling window, and $x(a, b)$ is the input value at position (a, b) within that window. This downsampling serves to make the representation approximately invariant to small translations of the input.

Dropout

To mitigate the overfitting, dropout layers are interspersed after fully connected layers, randomly setting a fraction p of input units to zero at each update during training time, which can be represented as: $D_i(x) = x \cdot \text{Bernoulli}(1 - p)$. Here, $\text{Bernoulli}(1 - p)$ is a vector of independent Bernoulli random variables taking the value 0 with probability p and 1 with probability $(1 - p)$.

Fully Connected Layers

The extracted features are flattened and passed through four fully connected layers, which focus the network's learning on relevant patterns for classification. The final output layer employs a sigmoid activation function to yield a probability indicating the presence of metastatic tissue: $\sigma(z) = \frac{1}{1+e^{-z}}$, where z is the input to the neurons in the output layer, and $\sigma(z)$ is the sigmoid function that squashes the output to be between 0 and 1, suitable for binary classification.

Training and evaluation

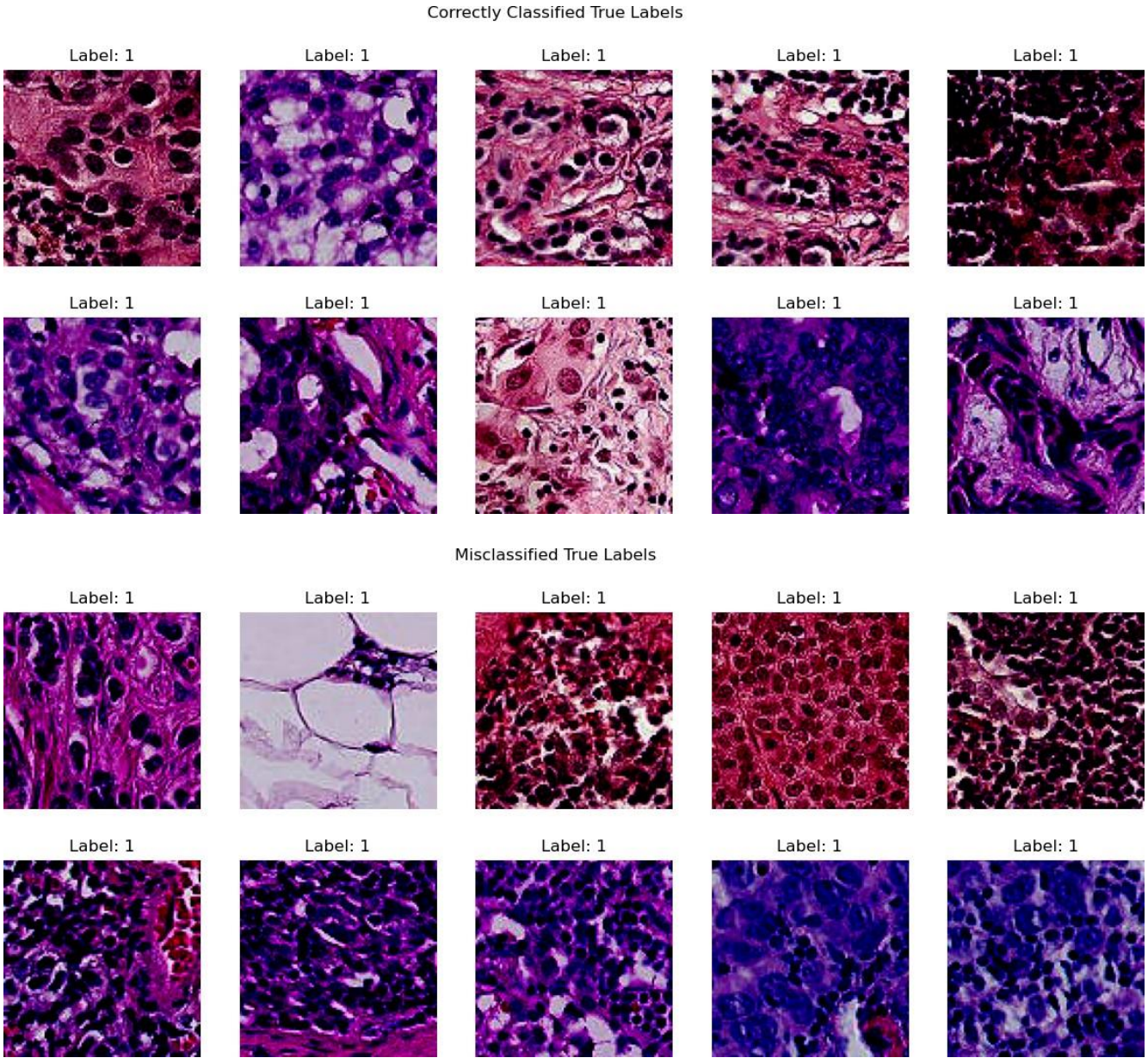
The model is trained for 50 epochs with a batch size of 64 on Nvidia A30 GPU (4 CPU cores, 16 GB memory). We used binary cross-entropy loss (BCEWithLogitsLoss) and the Adam optimizer, with a learning rate of 10^{-5}. We tracked several metrics during training, including accuracy, true positive rate (TPR), and true negative rate (TNR), providing insights into the model's performance. The model is evaluated on a separate validation set after each epoch and on another separate test set post training to assess the final

model's generalization capability and performance on unseen data. To avoid overfitting, we implement an early stopping mechanism. If the validation loss does not improve for a

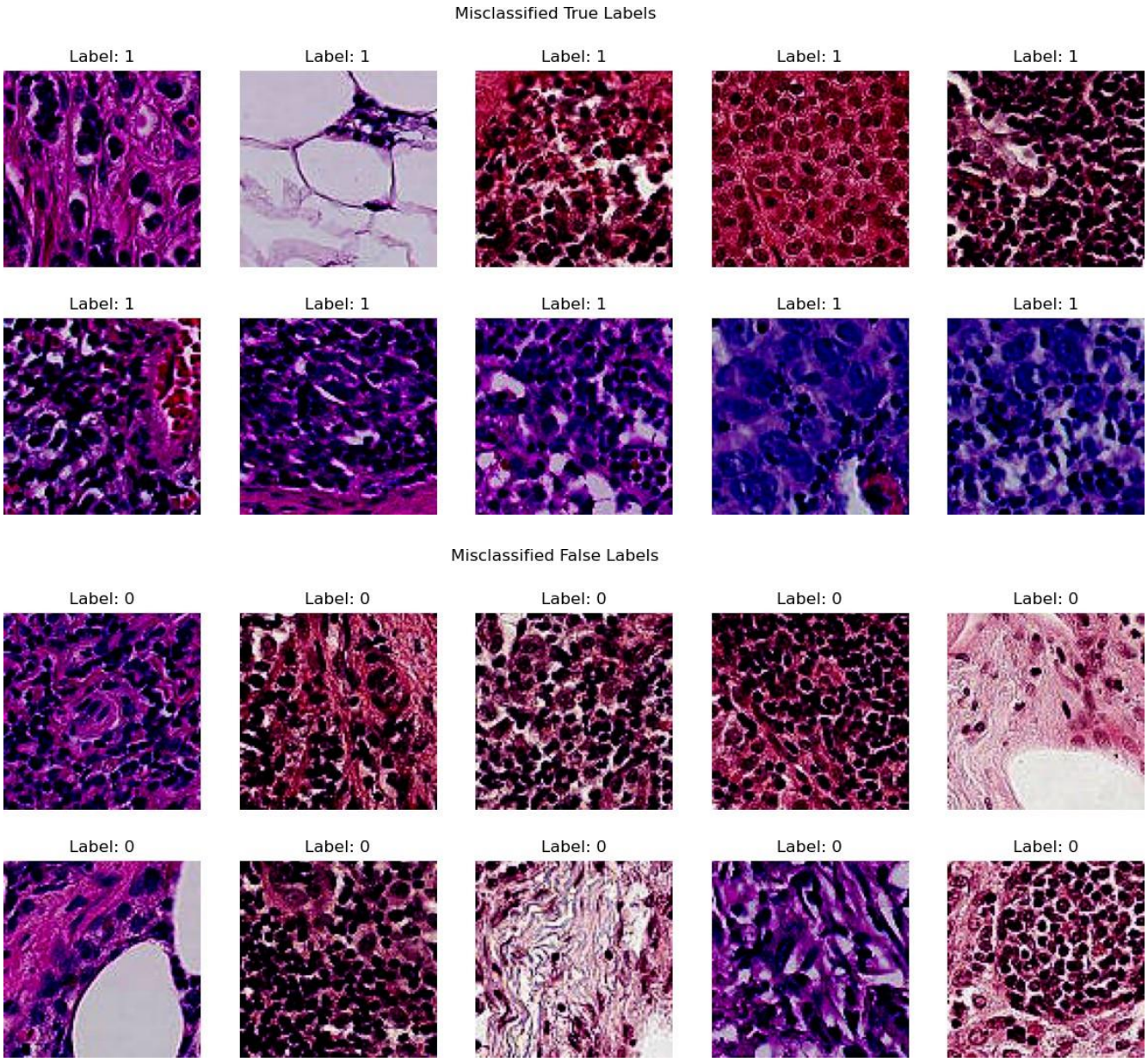
predetermined number of epochs (defined by 'early_stop_limit'), the training process is halted, ensuring that the model retains its generalization ability.

Result

Examples of correctly classified patches



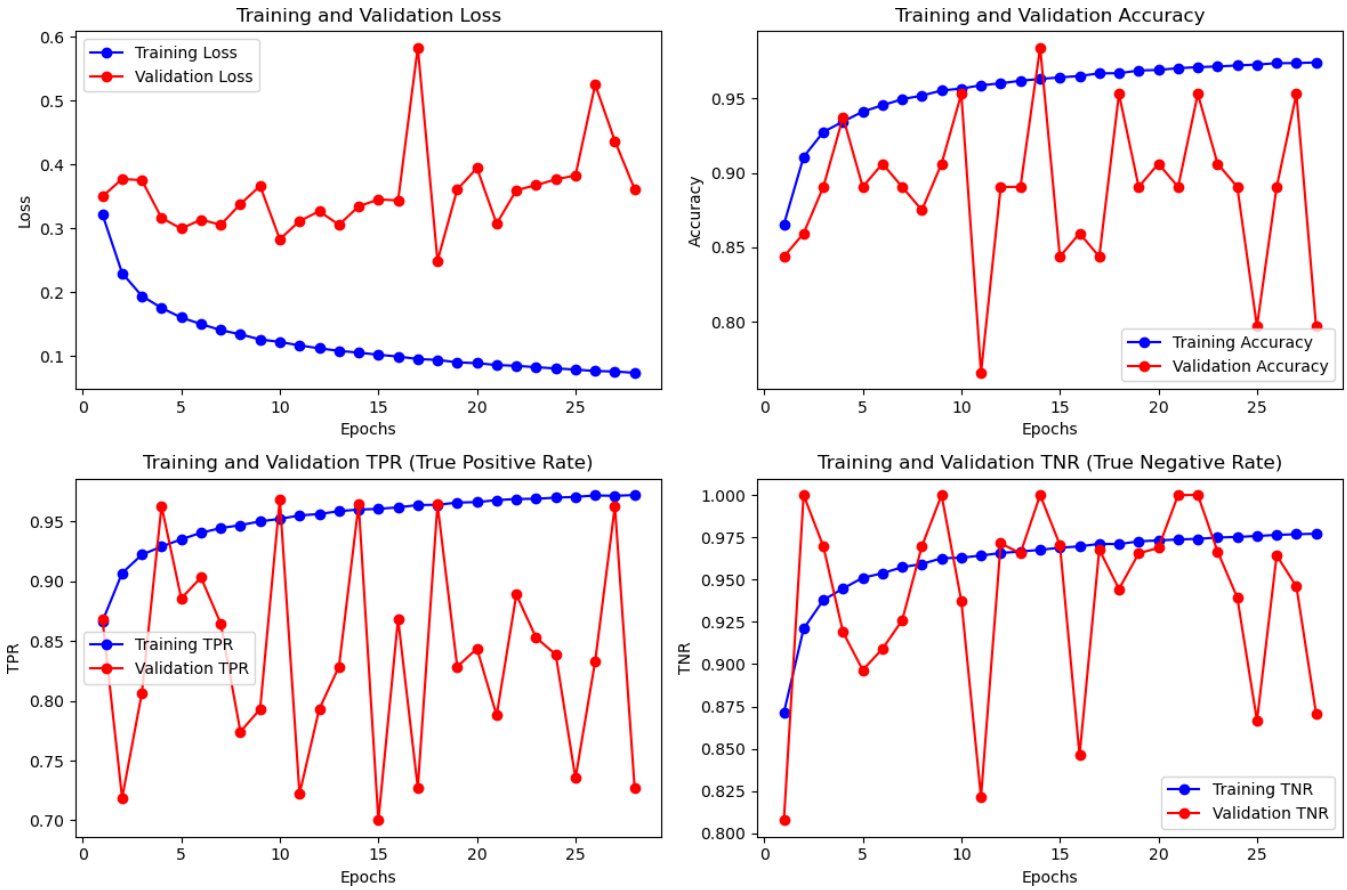
Examples of misclassified patches



Performance Metrics

Our model achieved an accuracy of 82.79%, with a precision of 85.28%, a recall of 82.79%, and an F1-score of 82.48% on the test dataset.

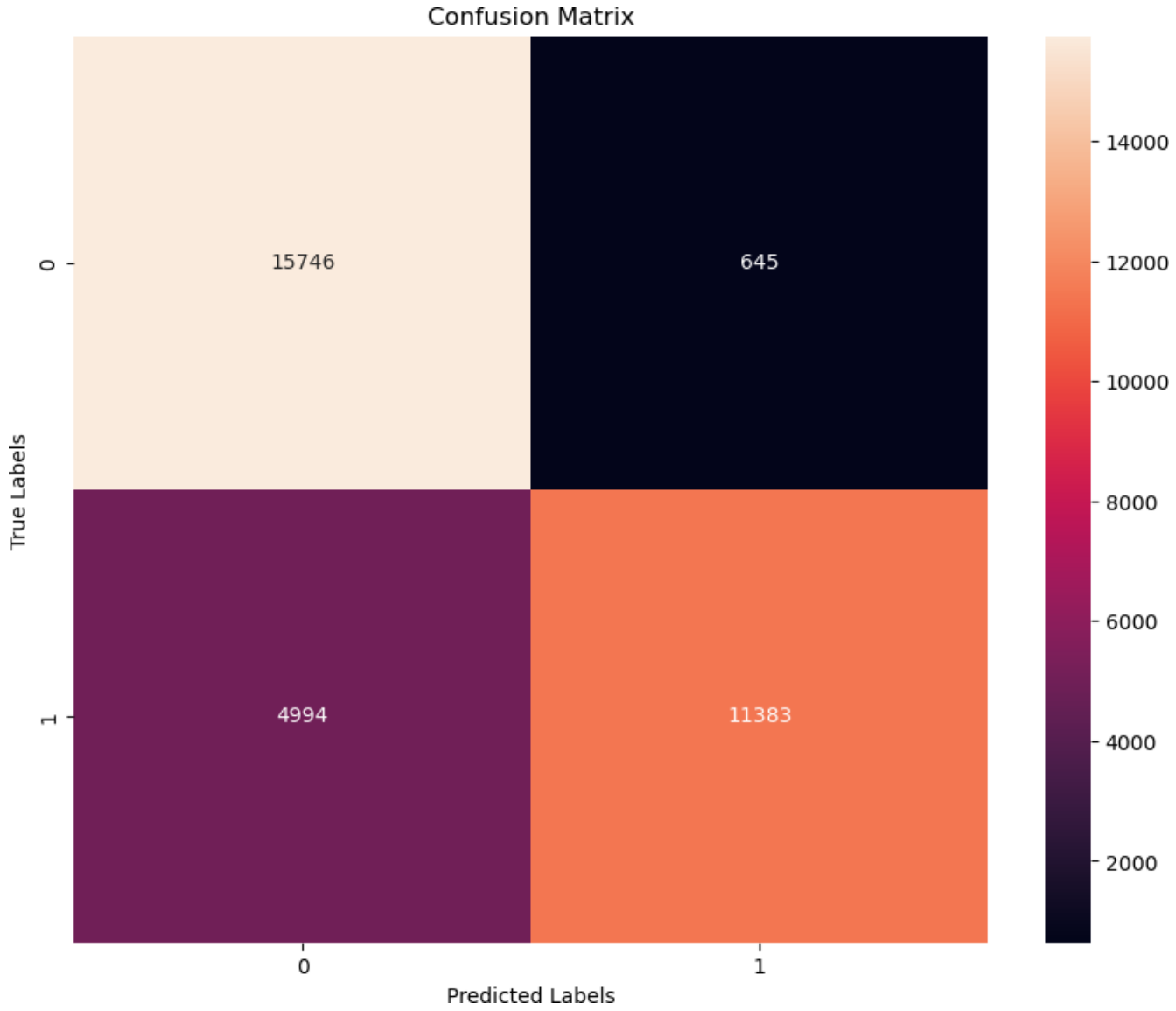
Training and Validation Trends



The analysis of training and validation loss indicates that the model is learning as expected. The training loss shows a general downward trend, indicating that the model is becoming better at predicting the training data over time. Validation loss, although more volatile, does not show signs of increasing, which suggests that the model is not overfitting to the training data.

The trends in accuracy and the rates of true positives (TPR) and true negatives (TNR) further reinforce the model's learning capability. The TPR remained consistently high, indicating the model's effectiveness in detecting true cancer cases. Similarly, the high TNR suggests that the model is also accurate in identifying true non-cancer cases. The accuracy remains high throughout the training epochs, with minor fluctuations due to the variability inherent in the validation dataset.

Confusion Matrix



The confusion matrix provides a visual and quantitative representation of the model's predictive capabilities. In our case, the matrix shows a relatively balanced distribution of predictions across the four quadrants. However, there is a notable number of false negatives (4994), which is critical in a medical context as it represents cases where the model missed detecting cancer.

Discussion

In this study, we trained a CNN model on the PCam dataset that demonstrated promising results. The architecture is constructed with the aim of achieving a balance between model complexity and computational efficiency, ensuring that it is capable of capturing essential features from the medical images without an excessive computational burden, thus rendering it suitable for use in a clinical setting. Overall, the model demonstrates promising results in identifying metastatic cancer in histopathologic images. The high

precision is particularly notable as it implies a low rate of false positives, which is essential to avoid unnecessary medical procedures. The achieved results are indicative of the model's potential as a tool to assist pathologists in diagnosing breast cancer. The use of CNNs in this context has proven valuable, and with further refinement, the model's reliability can improve, making it a more robust tool for medical diagnostics.

By accurately classifying patches of histopathologic scans, machine learning models can potentially assist pathologists in this task, reducing workload and improving diagnostic accuracy. In recent studies, models trained on the PCam dataset have demonstrated significant potential and sometimes exceed the performance of human experts in this specific task [3, 5]. The PCam dataset, with its well-defined problem space and large number of samples, continues to be an invaluable resource for researchers and practitioners aiming to advance the field of medical image analysis.

Despite the positive results, several challenges were noted during the training and evaluation phases. The number of false negatives suggests there is room for improvement, as these represent potentially missed diagnoses. The variability in the validation metrics suggests that the model might benefit from further training or data augmentation to stabilize its performance. Additionally, exploring other architectures or hyperparameter tuning could lead to improvements, particularly in reducing false negatives. In addition, the class imbalance present in the PCam dataset required careful handling to prevent the model from being biased towards the majority class. Furthermore, the high variability in the histopathological images meant that the model had to learn a vast array of features to make accurate predictions, which increased the computational complexity.

Moreover, while the model has demonstrated robustness in the contained environment of the dataset, its performance in a real-world clinical setting remains to be validated. Factors such as different staining protocols, image acquisition conditions, and inter-observer variability in slide annotations are not accounted for in the current training process and could affect the model's generalizability.

Overall, the CNN model demonstrates a high potential for aiding in the automated diagnosis of metastatic cancer in histopathological images. However, further work is needed to address its limitations and validate its performance in a clinical setting. Future research should focus on expanding the training dataset, implementing more sophisticated data augmentation techniques, and exploring more complex architectures or ensemble methods to improve the model's generalizability and robustness.

References

1. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Laak JA, Hermsen M, Manson QF, Balkenhol M, Geessink O. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama. 2017 Dec 12;318(22):2199-210.
2. Khened M, Kori A, Rajkumar H, Krishnamurthi G, Srinivasan B. A generalized deep learning framework for whole-slide image segmentation and analysis. Scientific reports. 2021 Jun 2;11(1):11579.
3. Wang X, Chen Y, Gao Y, Zhang H, Guan Z, Dong Z, Zheng Y, Jiang J, Yang H, Wang L, Huang X. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. Nature communications. 2021 Mar 12;12(1):1637.
4. Bandi P, Geessink O, Manson Q, Van Dijk M, Balkenhol M, Hermsen M, Bejnordi BE, Lee B, Paeng K, Zhong A, Li Q. From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. IEEE transactions on medical imaging. 2018 Aug 26;38(2):550-60.
5. Veeling BS, Linmans J, Winkens J, Cohen T, Welling M. Rotation equivariant CNNs for digital pathology. InMedical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11 2018 (pp. 210-218). Springer International Publishing.