**Pneumonia Image Classification: A Comparative Study of Naive Bayes and Support Vector Machine Models**

**Mike Rasell Carale Dago-oc1, Chris Angelu Bongcawil Jordan2**

Computer Science and Information Technology Department,

Bachelor of Science in Computer Science (BSCS)

Negros Oriental State University (NORSU), Negros Oriental, Philippines

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Keywords:***  Pneumonia  Naïve Bayes  Support Vector Machine (SVM)  Inception-V3  Accuracy  Precision  Recall  F1-Score |  | This study investigates the application of machine learning algorithms for the classification of chest X-ray images into Pneumonia and Normal categories. The dataset utilized comprises 5,863 anterior-posterior chest X-ray images collected from pediatric patients aged one to five years at the Guangzhou Women and Children’s Medical Center. Due to hardware limitations, a subset of 290 images (145 Pneumonia and 145 Normal) was used for this study. Pre-processing steps included image resizing, normalization, and conversion to meet the requirements of the Inception-V3 model for feature extraction. A comparative analysis was conducted between two machine learning algorithms: Naive Bayes (NB) and Support Vector Machine (SVM). The models were evaluated using 10-fold cross-validation. The results indicated that the SVM model outperformed NB, achieving an accuracy of 94.83%, with precision, recall, and F1-score values of 0.953, 0.948, and 0.948, respectively. The NB model also demonstrated strong performance, with an accuracy of 91.03% and precision, recall, and F1-score values of 0.919, 0.910, and 0.910, respectively. These findings suggest that the SVM model is highly effective for pneumonia classification in chest X-ray images, offering a reliable tool for automated diagnosis. |
|  |

1. **INTRODUCTION**

Pneumonia is one of the leading causes of death worldwide [1], it can affect everyone anywhere, especially children [2]. Pneumonia is a form of acute respiratory infection that is most commonly caused by viruses or bacteria targeting the lungs and it can cause mild to life-threatening illness in people of all ages. When an individual has pneumonia, the alveoli are filled with pus and fluid, which makes breathing painful and limits oxygen intake [3].

Identifying if a person has pneumonia requires general practitioners or doctors specializing in lung conditions [4]. Sadly some countries have not kept pace with the overall economic growth, and their health systems remain weak, which includes doctors who specialize in lung conditions [5]. This lack of general practitioners leads to increased workloads for existing specialists, potential delays in diagnosis and treatment, and overall strain on healthcare systems [6].

This study aims to make a machine-learning model that can classify X-ray images if it has pneumonia. With this tool, it can potentially automate diagnosis which can reduce the time doctors spend on diagnosing pneumonia cases. This allows doctors to focus more on treatment rather than performing or interpreting tests. Additionally, this can also be used as a training and education tool for medical students to learn how to interpret chest X-rays. However, the model is intended solely as a support tool, not a replacement for human judgment.

The rest of this paper is structured thus: Section 2 provides a detailed review of recent literature in the field of pneumonia image classification, Section 3 outlines the methodology employed in this study, and further provides a brief description of the machine learning algorithms used in this investigation. While Section 4 presents the results and discussion, and Section 5 concludes the research.

1. **LITERATURE REVIEW**

Naive Bayes (NB) and Support Vector Machine (SVM) have also been applied in image classification tasks due to their effectiveness in handling medical datasets. This review explores their applications in pneumonia detection using chest X-ray images, with an emphasis on methodologies and performance metrics.

**2.1. Image Classification**

One study tested the predictive performance of 6 machine learning models. In this review, we focus on Naive Bayes (NB) and Support Vector Machine (SVM). The other models include Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), and AdaBoost (AB). Their dataset consists of 5,856 X-ray images, which include 1,583 normal cases and 4,273 pneumonia cases and is split into 70% for training and 30% for testing. For image embedding they used 5 Deep Convolutional Neural Networks (DCNN) as feature extractors. However, this study will only focus on Inception-V3, the same as the one used in this study. The other DCNNs are AlexNet, SqueezeNet, VGG16, and VGG19. ANN delivered the best overall performance with a 97.19% accuracy, closely followed by LR at 97.08%. Meanwhile, SVM with a linear kernel achieved an accuracy of 85.02% [9].

Another study used five models including, Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), and Softmax. In image embedding they extracted the features from the fully connected (FC) layer of a modified CNN architecture using transfer learning. The FC layer generated feature vectors representing each class. The CNN model used in this study was pre-trained on Optical Coherence Tomography (OCT) images before being adapted to the pneumonia detection task. This differs from directly using models like Inception V3. Their dataset is comprised of 5,852 anterior-posterior chest X-ray images, which include 1,581 normal and 4,271 pneumonia cases, and is split into 70% for training, 15% for validation, and 15% for testing. However, in this review, we focus on the SVM, which achieved an accuracy of 99.61%. Among the models, Softmax slightly outperformed SVM with an accuracy of 99.72% [10].

Other researchers also used three Deep Convolutional Neural Networks (DCNN), which are VGG16, VGG19, and Inception-ResNet V2 in both extracting features and classifying the images after adding the DCNN classifier layers. They used datasets from three different sources which are from: Karolinska Directed Emotional Faces (KDEF), Extended Cohn-Kanade (CK+), and Net Images, which in total amount to 5235 images, which were then split to 80% for training, 10% for validation, and 10% testing. For feature extraction of the images, they used all three pre-trained models then they tested and added two types of classification layers for the final prediction of the image's class which are, Global Average Pooling (GAP) Classifier and Convolutional-Layer-Based Classifier. After fine-tuning, VGG16 with a convolutional layer-based classifier was the best-performing model with an overall accuracy of 89.6% and an f1-score of 89.7% [12].

1. **METHODOLOGY**

**3.1. Hardware and Software**

The study was carried out on a system running Windows 10 with a 64-bit operating system. The system uses an Intel Core™ i5-7400 CPU with 16GB Random Access Memory (RAM). The researchers utilized Jupyter and Python version 3.9.15 as the primary programming language for data analysis and model implementation with the following libraries: Math, NumPy, Pandas, PIL, OS, Scikit-Learn, Matplotlib, Seaborn, TensorFlow, and Keras.

**3.2. Data Acquisition**

This study employs a dataset from Kaggle, which is the Chest X-Ray Images (Pneumonia) dataset, a file with 5856 jpeg files which are anterior-posterior chest X-ray images. The images are carefully chosen from retrospective pediatric patients with age groups between 1 and 5 years from Guangzhou Women and Children’s Medical Center. The image has 2 normal categories pneumonia and normal which are further divided for testing, training, and evaluation [12]. Because of hardware limitations, this study uses 290 images with 145 normal and 145 pneumonia images.

**3.3. Image Embedding**

The main goal and objectives of the proposed system are to diagnose and make a tool that can identify whether a person has pneumonia through chest X-ray images. This study uses a Deep Convolutional Neural Network (DCNN) which is Inception-V3, for feature extraction of the X-ray images.

Inception models were developed for the first time in 2014. The structures of inception models and the conventional CNN model are different because they are inception blocks which means lapping the same input tensor with multiple filters and concatenating their results. In 2015, a researcher [15] proposed a new version of the inception models named Inception-V3, an improved version of the previous versions of inception models which are Inception-V1 and Inception-V2, and possesses 24M parameters. Inception-V3 improves the efficiency and performance of convolutional neural networks by introducing clever factorization techniques. Instead of directly using large convolutions (e.g., n x n), it breaks them into smaller, more manageable operations. For example, a 5 x 5 convolution is replaced by two 3 x 3 convolutions, and an n x n convolution is split into asymmetric 1 x n and n x 1 convolutions, reducing computation. Additionally, 7 x 7 convolutions are replaced with multiple 3 x 3 convolutions. Each inception block processes the input in parallel through multiple convolutional filters of different sizes (1 x 1, 3 x 3, and 5 x 5), as well as 3 x 3 max pooling, capturing features at various scales. These outputs are then concatenated and passed to the next module, allowing the network to learn complex features efficiently [13].

**3.4. Data Pre-processing**

Data processing is an essential aspect of model development. Data acquired in their raw form contain noise and anomalies, which can affect the performance and training process of the model being schooled [15]. In addition to this when preparing input data for a DCNN, it's essential to adhere to specific requirements regarding input shape, color channels, and preprocessing steps to ensure optimal performance [16]. The researchers employed several techniques to clean the data, which included image pre-processing and data normalization:

Image pre-processing is an essential part of image classification to ensure that input images align with the model's expectations. Inception-V3 requires input images in 299x299 resolution and should also have three color channels (red, green, and blue) [17][18]. And since the datasets used are X-ray images they only have 1 color channel (black and white) and inconsistent high resolution. With this in mind, the images are turned to have three color channels, and it is also resized to have a resolution of 299x299. After this, the image is preprocessed using TensorFlow Keras API preprocess\_input so that the images align with the model's training conditions.

Data normalization is a pre-processing technique primarily intended to manage numerical features and is applied to numerical features before the application of classification algorithms. Normalization is crucial to prevent the effect of certain features from being concealed by others, particularly when the ranges of the features are inconsistent [19]. After feature extraction using Inception-V3, the features are normalized to improve model training [20]. The normalization techniques used are Min-Max Normalization for Naive Bayes (NB), and Z-Score Normalization for the Support Vector Machine (SVM).

**3.5. Machine Learning Algorithms**

This section focuses on the machine learning classification models utilized in this study. After image and data pre-processing, the machine learning workflow progresses to the model training stage, where an algorithm is taught to learn from data and produce predictions. This algorithm is specifically responsible for the classification of whether an X-ray image has pneumonia. To find the best classifier for engagement level prediction, several classifiers, including Naive Bayes (NB) and Support Vector Machine (SVM) were tested through a variety of tests.

**3.5.1. Naïve Bayes**

The Naïve Bayes algorithm has its foundation rooted in the Bayes theorem by Thomas Bayes. One of the strengths of this model is its ability to handle missing values. And Unlike other models, Naïve Bayes conserves processing and training time [21][7]. The term ‘naive’ is used due to this algorithm's uncertain independence. With this, researchers in [22] stated that with this ability it's able to converge quicker when compared to several others.

(1)

Where P is the probability, X is the training set of attributes and Y is the given class [7].

**3.5.2. Support Vector Machine**

Support Vector Machine (SVM) is a binary linear classifier. As a non-probabilistic supervised learning algorithm, it utilizes training data and employs a high-dimensional space to construct a set of hyperplanes for data classification. While only the features of test data are provided, the model is trained on the training data to predict the target values. For effective classification of problem instances, SVM relies on selecting the optimal hyperplane [7].

**3.6 Model Evaluation Metrics:**

Evaluation measures are metrics used to assess the results of an experiment [24]. In the context of classification models, different evaluation metrics are used to measure their output. In this study, the main performance evaluation metric is “Accuracy”. However, additional metrics such as recall, precision, f-measure, and confusion matrices are also used to supplement the evaluation of the model's performance. Each model identifies learner engagement levels when assessed using these metrics. A brief description of these metrics is provided below.

Accuracy is a common evaluation metric for classification models. It's calculated as the ratio of well-predicted samples to the total sample of prediction. For a balanced dataset, accuracy is a reliable measure of the model's performance [24].

(2)

Where the numerator reflects the total number of correct predictions, while the denominator represents the total number of predictions made by the model. A higher accuracy value suggests that the model is more effective at correctly classifying both classes [7].

Precision measures the proportion of correctly predicted positive observations out of all the predicted positive observations. A high precision score indicates strong class predictions, while a low precision score reflects weak class predictions [24].

(3)

Where True Positives refer to the instances that were correctly identified as positive, while False Positives are the instances where the model wrongly predicted the positive class [7].

Recall is the ratio of correctly predicted positive results to all actual positive samples, also known as the detection rate. It's calculated by dividing the true positive samples by the sum of the positive samples [24].

(4)

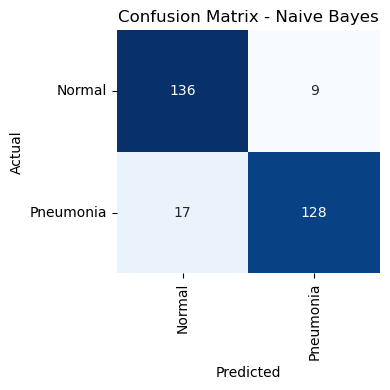
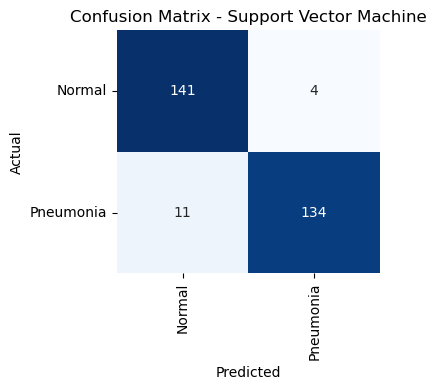
Where False Negatives occur when the model mistakenly classifies a positive instance as belonging to the negative class [7].

F1-score is the mean value for recall and precision. It offers an indicator of mistakenly graded results [24]. It is regarded as the best metric for measuring the performance of models on an imbalanced dataset. It ranges from 0 to 1, with higher values indicating better model performance [24].

(5)

Where Precision evaluates the accuracy of positive predictions and Recall assesses the model's ability to identify all relevant instances, the F1-Score combines these two metrics into a single value, offering a balanced measure of the model's accuracy and completeness [7].

1. **RESULTS AND DISCUSSION**

**4.1. Confusion Matrix**

**Figure 2.** SVM Confusion Matrix

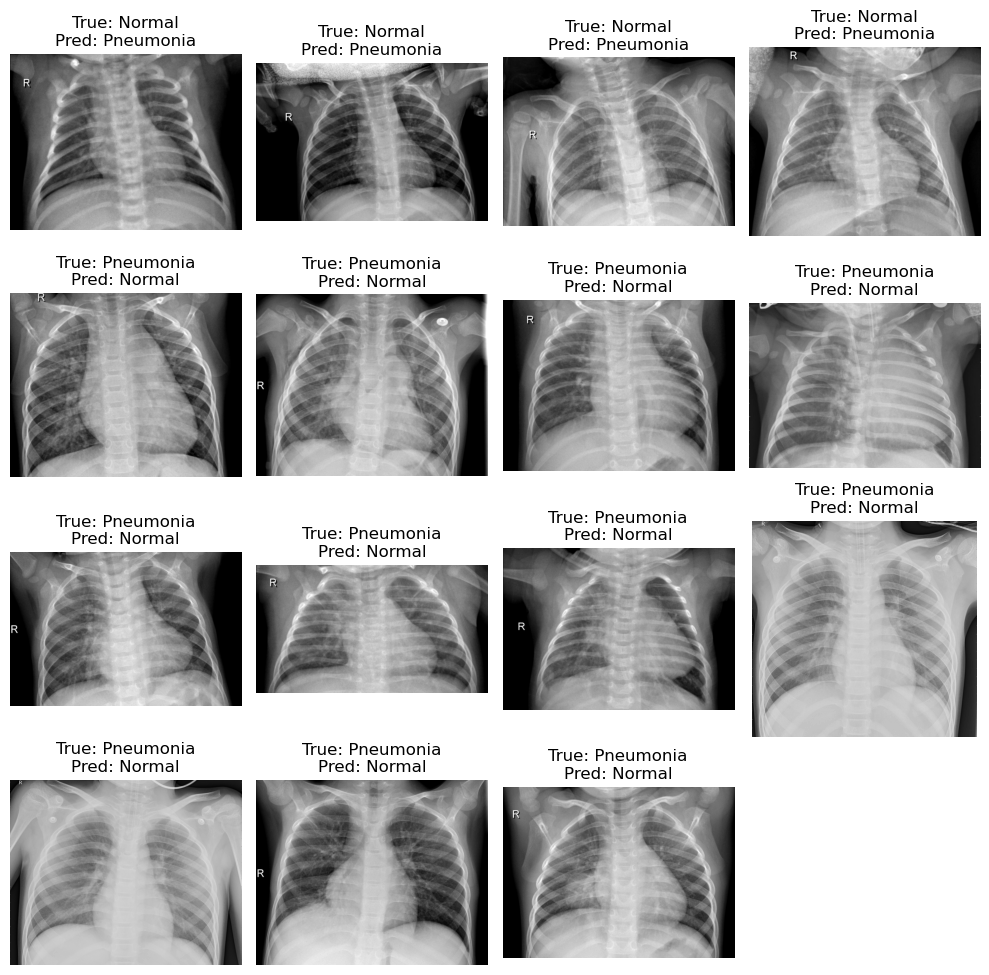
**Figure 3.** NB Confusion Matrix

**Figure 1.** SVM Confusion Matrix

**Figure 2.** NB Confusion Matrix

In Figure 1, the confusion matrix for the SV M model shows high accuracy in detecting both “Normal” and “Pneumonia” cases. With only four false negatives, which is critical for pneumonia detection since missing a pneumonia case could have severe consequences. Overall, SVM shows slightly better overall performance compared to Naive Bayes in terms of reducing misclassification.

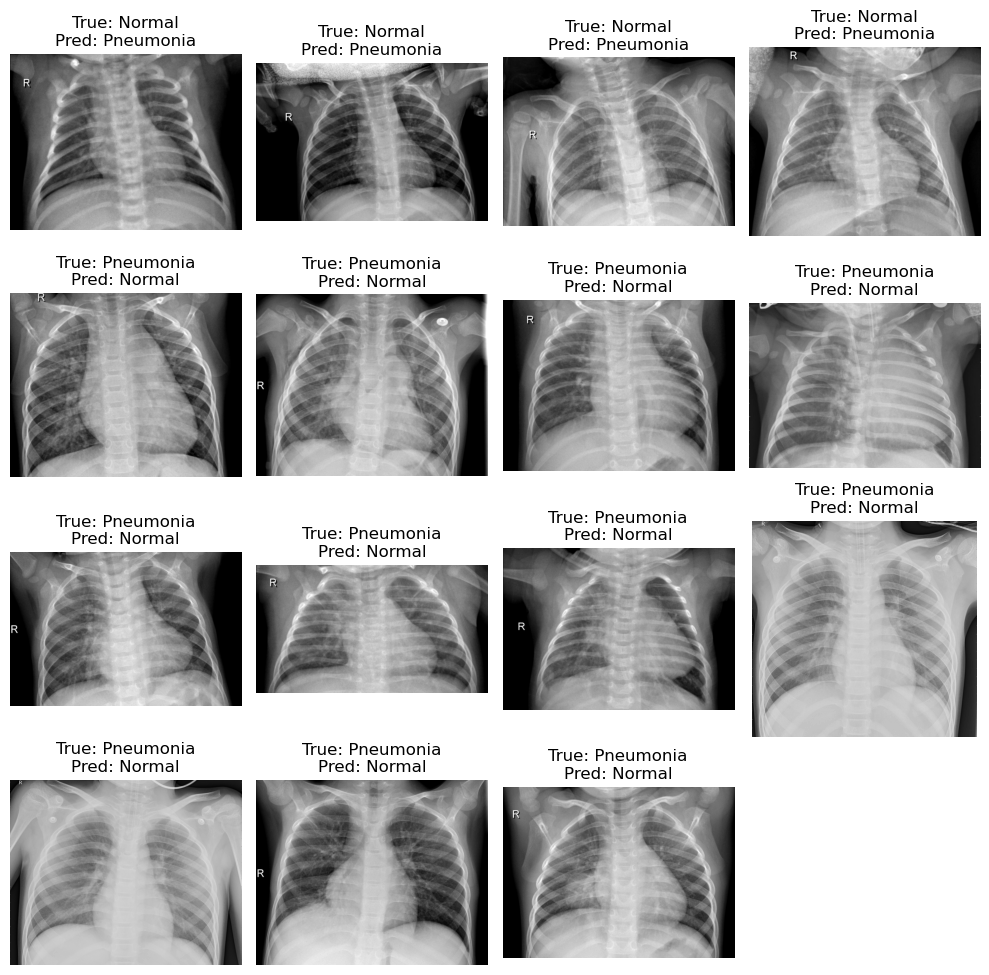
In contrast in Figure 2, NB shows slightly lower accuracy compared to SVM, as evidenced by higher false positives (17) and false negatives (9). Naive Bayes struggles more with classifying pneumonia cases accurately which results in higher false negatives, which is a critical area for improvement. This model may be more prone to noise or incorrect feature distribution assumptions, which could explain the performance gap.

**4.2 Misclassified Image Analysis**

**Figure 3.** SVM Misclassified Images

The first row shows chest X-rays where the true label is “Normal”, but the SVM model predicted “Pneumonia”. This suggests slight opacities or irregularities in the lung regions of the misclassified chest X-ray images that resemble pneumonia patterns. The rest of the rows are chest X-rays where the true label is ”Pneumonia”, but the model predicted “Normal”. This indicates that the misclassified chest X-rays may lack clear and distinct patterns of pneumonia.

These misclassifications occur in cases where the X-rays are visually less distinct. The model also might be struggling with nuanced or borderline cases where the features extracted by Inception V3 are less pronounced.



**Figure 4.** NB Misclassified Images

The first and second half rows highlight cases where the model predicted “Normal” as a “Pneumonia” image. The rest of the rows showcase where “Pneumonia” images were predicted as a “Normal” image. It seems that the misclassifications occur for borderline or vague cases which is the same as the SVM model. The NB model might struggle with complex patterns due to its simplicity and assumptions of feature independence, which can lead to limitations in capturing nuanced or overlapping features in X-ray images.

**4.3. Model Performance**

**Table 1**. Performance Metrics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1-Score |
| SVM | 0.9483 | 0.9528 | 0.9483 | 0.9480 |
| NB | 0.9103 | 0.9186 | 0.9103 | 0.9095 |

In the evaluation of the two machine learning models using a 10-fold cross-validation, in Table 1 it can be observed that the performance of all the models accuracy ranges from 90% to 94%. The result shows that all the models can strongly identify whether X-ray images have pneumonia. And with a thorough analysis of the results, indicates that SVM provided the highest accuracy of 94.83%, with a precision of 0.9528, a recall of 0.9483, and an F1-score of 0.9480. The NB model, while performing well, showed slightly lower accuracy compared to SVM, with an accuracy of 91.03%, a precision of 0.9186, a recall of 0.9103, and an F1-score of 0.9095. These results suggest that SVM handles the dataset very well, using kernel functions like the Radial Basis Function (RBF) shows that SVM is a powerful model for this study. While NB performs relatively well, its performance is lower than the other models. Overall, the findings highlighted the effectiveness of all two models and emphasized their potential for application in automated pneumonia detection.

1. **CONCLUSION**

In conclusion, the SVM model demonstrated superior performance in classifying chest X-ray images for pneumonia detection, achieving the highest accuracy of 94.83%, along with strong precision, recall, and F1-score metrics. The SVM model consistently excelled in distinguishing between Pneumonia and Normal cases, showcasing its robustness in handling complex feature interactions and high-dimensional data. However, minor misclassifications were observed, particularly in borderline cases, where subtle or less distinct patterns in the X-rays may have led to false negatives or positives.

The NB model also delivered strong results, achieving an accuracy of 91.03%. While slightly less effective than SVM, it proved reliable for simpler patterns in the data. However, its assumption of feature independence limited its ability to handle overlapping or nuanced features, resulting in higher false negatives and false positives compared to SVM. This suggests that NB may benefit from additional preprocessing or feature engineering to enhance its performance.

Overall, the SVM model emerged as the more effective classifier for pneumonia detection, particularly in applications where minimizing false negatives is critical. Nevertheless, both models underscore the potential of machine learning techniques in automating the diagnosis of pneumonia, offering valuable tools to complement clinical decision-making and improve healthcare outcomes.

**REFERENCES:**

[1] Our World in Data, "Pneumonia," 2025. [Online]. Available: https://ourworldindata.org/pneumonia. [Accessed: 12-Jan-2025].

[2] World Health Organization, "Pneumonia," WHO, Dec. 11, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia. [Accessed: Jan. 12, 2025]

[3] World Health Organization, "Pneumonia," World Health Organization, [Online]. Available: https://www.who.int/health-topics/pneumonia#tab=tab\_1. [Accessed: Jan. 12, 2025]

[4] Mayo Clinic, "Pneumonia - Diagnosis and treatment," Mayo Clinic, Dec. 15, 2022. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/pneumonia/diagnosis-treatment/drc-20354210. [Accessed: Jan. 12, 2025].

[5] J. M. Chakaya, E. J. Carter, and P. C. Hopewell, "Pulmonary specialty training to improve respiratory health in low- and middle-income countries. Needs and challenges," Annals of the American Thoracic Society, vol. 12, no. 4, pp. 486–490, Apr. 2015, doi: 10.1513/AnnalsATS.201502-071PS.

[6] P. Mishra, S. Bag, and S. Nandi, "AI-driven computational analysis of COVID-19 genome reveals distinct association patterns," Journal of Biomedical Research & Innovation, vol. 13, no. 6, pp. 112–127, Dec. 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9799366/

[7] G. Theophilus and C. I. Eke, "Machine learning-based e-learners' engagement level prediction using benchmark datasets," Int. J. Appl. Inf. Syst., vol. 12, no. 41, Sept. 2023.

[8] EIT Health, "Machine learning in healthcare: Uses, benefits and pioneers in the field," EIT Health, 18-Sep-2024. [Online]. Available: https://eithealth.eu/news-article/machine-learning-in-healthcare-uses-benefits-and-pioneers-in-the-field/. [Accessed: 12-Jan-2025].

[9] Y. Muhammad, M. D. Alshehri, W. M. Alenazy, T. V. Hoang, and R. Alturki, "Identification of pneumonia disease applying an intelligent computational framework based on deep learning and machine learning techniques," Mobile Inf. Syst., vol. 2021, Art. no. 9989237, 20 pages, May 18, 2021. [Online]. Available: https://doi.org/10.1155/2021/9989237. [Accessed: 12-Jan-2025].

[10] I. S. Masad, A. Alqudah, A. M. Alqudah, and S. Almashaqbeh, "A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest X-ray images," Int. J. Electr. Comput. Eng. (IJECE), vol. 11, no. 6, pp. 5530-5540, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5530-5540.

[11] T.B. Chandra and K. Verma, "Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm," in Proceedings of 3rd International Conference on Computer Vision and Image Processing, B. Chaudhuri, M. Nakagawa, P. Khanna, and S. Kumar, Eds., vol. 1022, Advances in Intelligent Systems and Computing, Springer, Singapore, 2020, pp. 15-22. doi: 10.1007/978-981-32-9088-4\_3.

[12] P. T. Mooney, Chest X-ray Images (Pneumonia), Kaggle. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data. [Accessed: 12-Jan-2025].

[13] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” Technical report, Cornell University, New York, NY, USA, 2014.

[14] M. C. M. Pasague, "Classifying Garbage Images: A Comparative Study of Naive Bayes, SVM, and Logistic Regression Models," Bachelor of Science in Computer Science, Negros Oriental State University, 2025. [Online]. Available: cristinapasague27@gmail.com.

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, “Rethinking the inception architecture for computer vision,” 2015, https:// arxiv.org/abs/1512.00567.

[16] P. Kragsterman, "The ultimate guide to preprocessing medical images: Techniques, tools, and best practices for enhanced diagnosis," About CMRAD, Aug. 28, 2024. [Online]. Available: https://about.cmrad.com/articles/the-ultimate-guide-to-preprocessing-medical-images-techniques-tools-and-best-practices-for-enhanced-diagnosis. [Accessed: Jan. 12, 2025].

[17] Keras, "InceptionV3," Keras API Documentation. [Online]. Available: https://keras.io/api/applications/inceptionv3/. [Accessed: Jan. 12, 2025].

[18] Pytorch Team, "Inception v3," \*PyTorch\*, 2021. [Online]. Available: https://pytorch.org/hub/pytorch\_vision\_inception\_v3/. [Accessed: Jan. 12, 2025].

[19] B. S. Oliveira, A. M. G. de Oliveira, and A. C. de Albuquerque, "The Effect of Preprocessing Techniques, Applied to Numeric Features, on Classification Algorithms’ Performance," Algorithms, vol. 6, no. 2, pp. 11, 2023. [Online]. Available: https://www.mdpi.com/2306-5729/6/2/11. [Accessed: Jan. 12, 2025].

[20] A. Choudhary, "Feature Scaling in Machine Learning: Normalization vs Standardization," Analytics Vidhya, Apr. 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/. [Accessed: Jan. 12, 2025].

[21] E. F. Buraimoh, “Predicting Student Success Using Student Engagement in the Online Component of a Blended-Learning Course,” pp. 1–65, 2021.

[22] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, “Student Performance Prediction Model based on Supervised Machine Learning Algorithms,” IOP Conf. Ser. Mater. Sci. Eng., vol. 928, no. 3, 2020, doi: 10.1088/1757- 899X/928/3/032019.

[23] P. Sun, M. D. Reid, and J. Zhou, “An improved multiclass LogitBoost using Adaptive-One-Vs-One,” Mach. Learn., vol. 97, no. 3, pp. 295–326, 2014, doi: 10.1007/s10994- 014-5434-3.

[24] C. I. Eke, A. A. Norman, and L. Shuib, Multi-feature fusion framework for sarcasm identification on twitter data: A machine learning based approach, vol. 16, no. 6

[25] H. Hassan, N. B. Ahmad, and S. Anuar, “Improved students’ performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining,” J. Phys. Conf. Ser., vol. 1529, no. 5, 2020, doi: 10.1088/1742-6596/1529/5/052041.