**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.

1. **LITERATURE REVIEW**

The application of machine learning algorithms, particularly Support Vector Machines (SVM) and Naïve Bayes (NB), has significantly advanced the field of medical image classification. Both algorithms are widely used for their effectiveness and adaptability in handling complex datasets. This literature review examines recent studies utilizing SVM and NB for image classification tasks, emphasizing their methodologies and outcomes in medical imaging and related fields.

**2.1. Image Classification**

Image classification has been a central focus of machine learning research, with numerous advancements driven by improved algorithms and computational techniques. Traditional machine learning methods, including Naïve Bayes and Support Vector Machines, have consistently demonstrated their efficacy in classifying medical images such as chest X-rays, MRI scans, and CT images.

**2.2. Naïve Bayes Algorithm**

Naïve Bayes, a probabilistic classifier based on Bayes' theorem, excels in computational efficiency and is effective for datasets with independent features. One study tested the predictive performance of six machine learning models, including Naïve Bayes, using a dataset of 5,856 chest X-ray images, which included 1,583 normal and 4,273 pneumonia cases. The dataset was split into 70% for training and 30% for testing. For feature extraction, the study used the Inception-V3 model along with other DCNNs such as AlexNet, SqueezeNet, and VGG19. Naïve Bayes achieved an accuracy of 91%, showcasing its ability to handle noise and imbalanced datasets effectively. However, NB’s assumption of feature independence can limit its performance in complex scenarios where feature interactions play a critical role [7].

**2.3. Support Vector Machine Algorithm**

On the other hand, Support Vector Machines are powerful supervised learning models capable of handling high-dimensional data. SVM with a linear kernel achieved an accuracy of 85.02%. Although not the highest among the tested models, this performance highlights SVM’s capacity for effective classification when paired with appropriate feature extraction techniques [7].

Another study explored the use of SVM in pneumonia classification, employing a modified CNN architecture with transfer learning for feature extraction. The dataset consisted of 5,852 chest X-ray images, split into 70% for training, 15% for validation, and 15% for testing. SVM achieved an accuracy of 99.61%, a significant improvement attributed to the integration of advanced feature extraction techniques. This study underscores SVM’s ability to generalize well in high-dimensional and complex datasets [8].

Additionally, researchers tested various DCNNs, including VGG16, VGG19, and Inception-ResNet V2, for feature extraction and classification. These networks were evaluated using datasets from diverse sources, totaling 5,235 images. For classification, SVM combined with DCNN-generated features achieved competitive performance, with accuracies surpassing 89.6% in certain configurations. These results highlight the importance of leveraging deep learning models for feature extraction to maximize the performance of traditional algorithms like SVM [9].

1. **METHODOLOGY**

**3.1 Materials**

**3.1.1 Datasets**

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 [10]. Because of hardware limitations, this study uses 290 images with 145 normal and 145 pneumonia images.

**3.1.2 Hardware**

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).

**3.1.3 Software**

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 Methods**

**3.2.1 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 [11]. 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 [12]. 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) [13][14]. 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 [15]. Normalization is crucial to prevent the effect of certain features from being concealed by others, particularly when the ranges of the features are inconsistent [16]. After feature extraction using Inception-V3, the features are normalized to improve model training [17]. The normalization techniques used are Min-Max Normalization for Naive Bayes (NB), and Z-Score Normalization for the Support Vector Machine (SVM).

**3.2.2 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.

**3.2.2.1 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 [18][19]. The term ‘naive’ is used due to this algorithm's uncertain independence. With this, researchers stated that with this ability it's able to converge quicker when compared to several others [20].

(1)

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

**3.2.2.2 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 [19].

The formula for the Support Vector Machine (SVM) decision boundary can be expressed as:

(2)

Where is the weight vector, determining the orientation of the hyperplane,  is the feature vector of the input data,  is the bias term, shifting the hyperplane

**3.2.3 Feature Extraction**

The main goal and objectives of the proposed system are to diagnose and create a tool that can identify whether a person has pneumonia through chest X-ray images. This study uses 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 convolutional neural networks (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 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 CNNs 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 [20].

**3.2.4 Model Evaluation Metrics:**

Evaluation measures are metrics used to assess the results of an experiment [21]. 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.

(3)

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 [21].

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.

(4)

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 [21].

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.

(5)

Where False Negatives occur when the model mistakenly classifies a positive instance as belonging to the negative class [24].

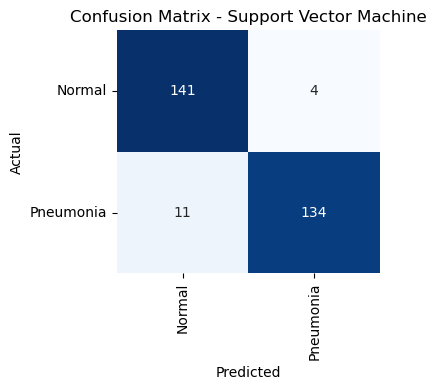
F1-score is the mean value for recall and precision. It offers an indicator of mistakenly graded results [24]. It is regarded as the best metric for measuring the performance of models on an imbalanced dataset. It ranges from 0 to 1, with higher values indicating better model performance.

(6)

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 [21].

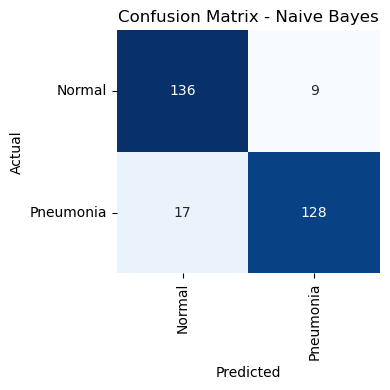
1. **RESULTS AND DISCUSSION**

**4.1. Confusion Matrix**



**Figure 1.** SVM Confusion Matrix

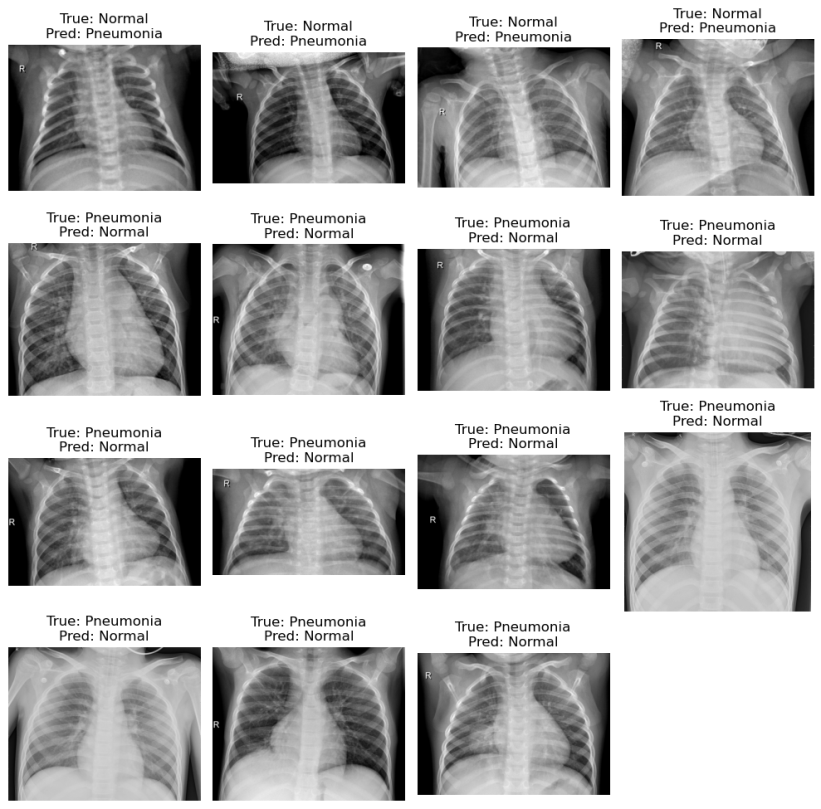
In Figure 1, the confusion matrix for the SV M model shows high accuracy in detecting both “Normal” and “Pneumonia” cases. With only four false negatives, which is critical for pneumonia detection since missing a pneumonia case could have severe consequences. Overall, SVM shows slightly better overall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



**Figure 2.** NB Confusion Matrix

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.

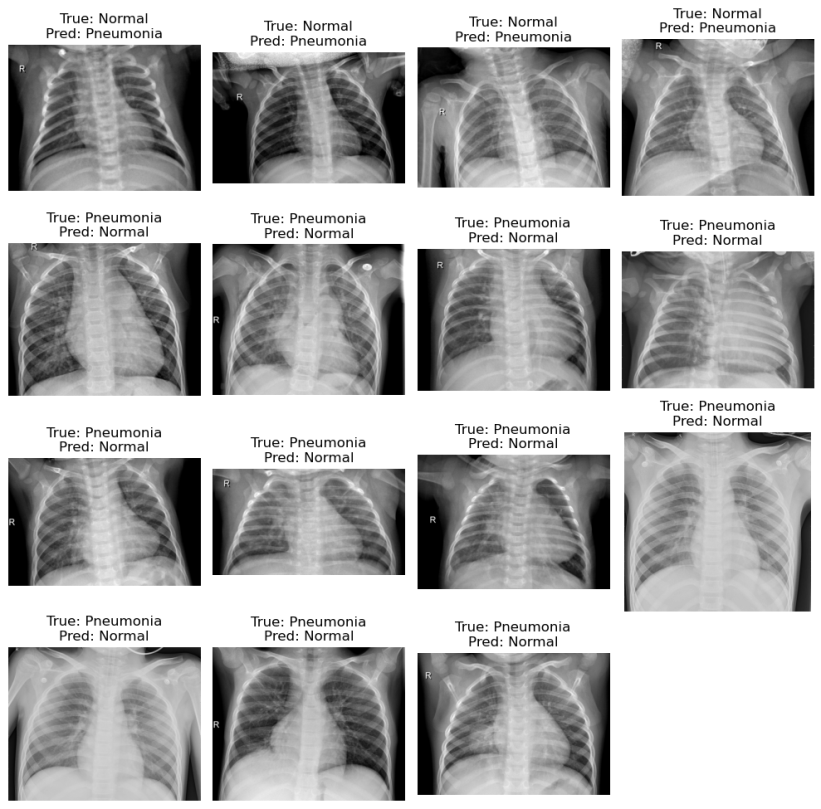
**4.2 Misclassified Image Analysis**



**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] 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].

[8] 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.

[9] 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.

[10] 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].

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

[12] 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].

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

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

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

[16] 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].

[17] 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].

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

[19] 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.

[20] 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.

[21] 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