WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Development of a Monkeypox Prediction Model Using Deep Learning and Grey Wolf Optimization (GWO) Algorithm

Nwobodo-Nzeribe Nnenna Harmony[a], Shamsudeen Mohammed S.[b], Njoku Camillus Ekene[c]

[a] nnennanwobo8@gmail.com [b] shamsudeensada@yahoo.com, [c]njokucamillus@gmail.com

[a, b, c] Department of Computer Engineering, Enugu State University of Science and Technology Enugu, Nigeria

ABSTRACT

This study presents the development of a Monkeypox prediction model using Convolutional Neural Network (CNN) optimized through the integration of Grey Wolf Optimization (GWO) algorithm. The dataset used for the implementation is a Kaggle dataset with 24,500 cases and 11 clinical features in order to create a reliable diagnostic method. Missing value elimination and data balancing using the SMOTEEN algorithm were preprocessing techniques that reduced class imbalance and ensured data quality. The CNN model was meticulously built utilising regularisation, max pooling, and Rectified Linear Unit (ReLU) activation techniques to prevent overfitting and underfitting. Essential parameters like as learning rate, batch size, and filter size were adjusted through hyperparameter optimisation utilising the GWO approach in order to get the best possible model performance. The proposed model demonstrated exceptional accuracy (95.21%), precision (95.54%), recall (98.05%), and F1-score (96.78%) with an AUC score of 92.59%, indicating its robustness in detecting monkeypox patients. This work demonstrates the potential of integrating optimisation techniques with deep learning for the early identification of infectious diseases. The proposed approach addresses significant challenges in medical diagnostics and provides a framework for developing accurate, reliable, and efficient diagnostic systems for public health applications.

Keywords: Monkeypox Prediction; CNN; Grey Wolf Optimization (GWO); Feature Selection; SMOTEEN algorithm

1. INTRODUCTION

A new viral disease known as Mpox has been detected in 75 countries and is currently causing a global outbreak that is spreading quickly (Hussain et al., 2022). A Danish laboratory employing monkeys discovered the first mpox virus outbreak in 1958 (Moore et al., 2022). In 1970, the virus was first known to spread from person to animal in the Democratic Republic of the Congo (Thornhill et al., 2022). When an outbreak of suspected mpox was detected in an 11-year-old child in 2018, the disease was considered serious. On January 1, 2019, there were 132 confirmed cases of this illness and seven deaths. The World Health Organisation (WHO) announced on June 23, 2022, that over 3000 mpox virus infections have occurred in over 50 nations in five regions (Thornhill et al., 2022).

This disease has drawn the attention of the WHO, which has deemed it a global health emergency (WHO, 2023). The difficulties in diagnosing mpox include the fact that symptoms are sometimes vague and that a diagnosis may not be made for days. Furthermore, the lack of widespread access to Polymerase Chain Reaction (PCR) testing hinders prompt disease diagnosis. To prevent the disease's transmission, mpox must be identified early. However, while mpox and other diseases including scarlet fever, roseola, and smallpox are similar, early identification is challenging (Jaradat et al., 2023).

Medical image analysis has recently attracted a lot of interest in artificial intelligence (AI) methods including machine learning and deep learning. These techniques have showed promise in a variety of applications, including the diagnosis of skin diseases. According to Hahn et al. (2021), Convolutional Neural Networks (CNNs) are a powerful tool for medical image analysis and have shown remarkable efficacy in image analysis tasks.

There are many difficulties with using CNNs to categorise skin lesions (Kattenborn et al., 2021). To begin with, training and validating CNN models is challenging due to the scarcity of big and high-quality datasets of monkeypox skin lesions. Second, CNN models may not accurately detect monkeypox lesions due to variations in their size, shape, colour, texture, and position on the body (Kawakami et al., 2019; Lin et al., 2021). Third, the CNN models may misclassify or confuse some lesions due to their overlapping characteristics with other skin disorders or diseases. Finally, to guarantee the CNN models' robustness, generalisability, and dependability in actual clinical situations, additional testing and validation on bigger and more varied datasets could be necessary (Keser et al., 2022).

This study's main goal is to use CNNs and the Grey Wolf Optimisation (GWO) approach to develop an accurate and reliable system for the automated classification of monkeypox skin lesions. Monkeypox skin lesions may be successfully classified to help with early diagnosis, treatment, and detection, which will eventually improve patient outcomes.

2. RESEARCH METHODOLOGY

This study's methodology discusses the Monkeypox prediction model, which consists of four stages which includes; pre-processing the data, feature selection from the pre-processed data to determine the most important symptoms that can improve the diagnosis's accuracy, Monkeypox prediction using the CNN model and CNN hyperparameter optimisation using the GWO algorithm. In the last step, the model is assessed using measures including F1-score, recall, accuracy, and precision. The model that performs the best is chosen to be the final model. The four stages of the suggested model are shown in Figure 1.

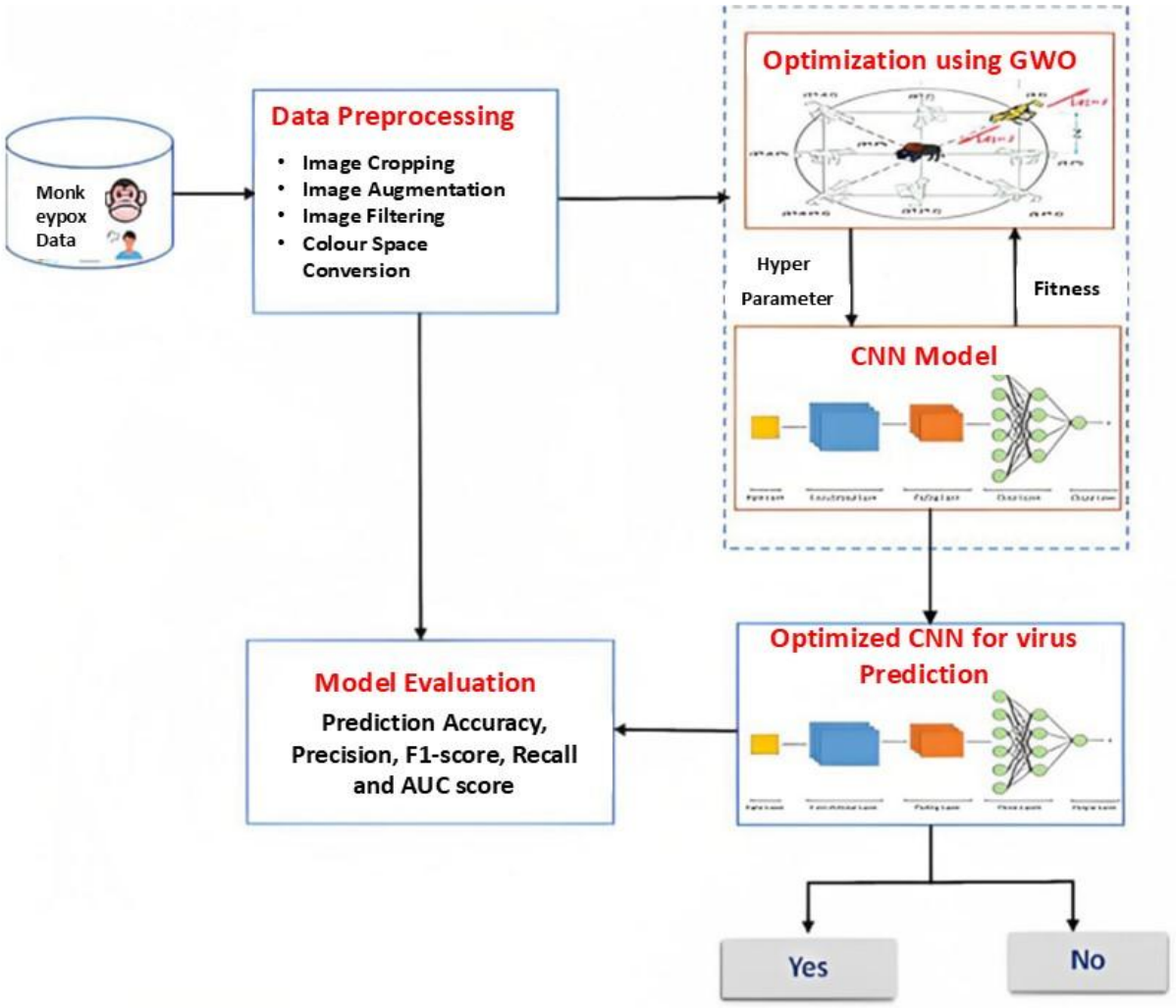

Figure 1: Methodology of the Proposed System using Optimized CNN for Monkeypox Prediction

2.1 Data Acquisition

The Monkeypox dataset is the one utilised to construct the suggested model. The dataset documents the clinical manifestations of human monkeypox infection during the epidemic at a major London location in 2022. The dataset, which includes 24,500 cases with 11 characteristics and a target variable indicating the presence or absence of monkeypox, is openly accessible on Kaggle and some of the symptoms considered in the dataset are fever, enlarged lymph nodes, aches and pains in the muscles, pains in the rectal area, sore throat, penile oedema, oral lesions, a single lesion, enlarged tonsils, HIV infection, and sexually transmitted infection. Sample photos of monkeypox data are shown in Figure 2.

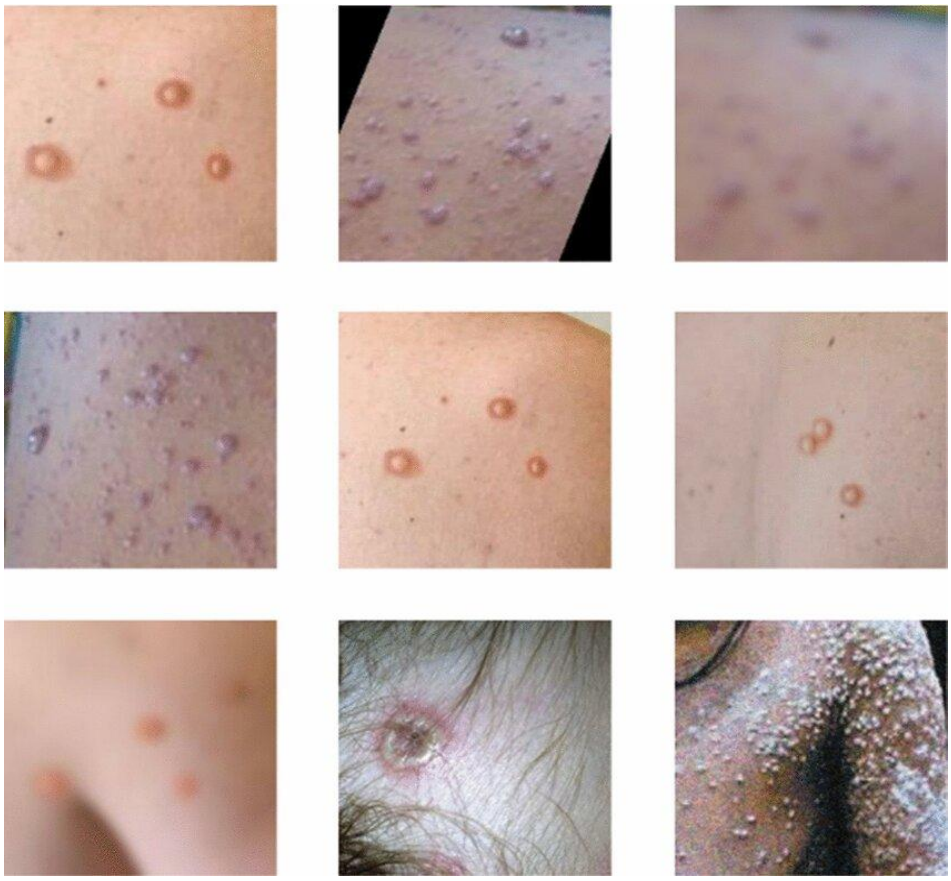

Figure 2: Sample of Monkeypox dataset (Gupta et al., 2023)

2.2 Data Preprocessing

The data must be cleaned and pre-processed to eliminate noise and extraneous information in order to increase the classification model's accuracy and efficiency. This process of data cleaning for the Monkeypox dataset removes missing values. SMOTEEN which is a combination of the Synthetic Minority Over-sampling Technique (SMOTE) (Elreedy et al., 2024) with Edited Nearest Neighbours (ENN) (Alejo et al., 2010) to produce a balanced dataset that is less prone to noise, is a widely used technique for data balancing (Muaz et al., 2020).

2.3 Feature Extraction

Selecting the most significant features is a crucial stage in creating prediction models as it has a direct impact on the models' performance, after preparation of the monkeypox dataset by removing null values and balancing the data using the SMOTEEN approach (Boukif et al., 2018). One useful technique for determining the connections between a dataset's properties is correlation analysis. The most significant characteristics may be chosen, and the number of variables in the model may be decreased, by determining which variables have a high association with the target variable. As a result, this can decrease overfitting and increase the model's accuracy. One method for showing the correlation coefficients between variable pairs in a dataset is to create a correlation matrix. Correlation coefficients, which range from -1 to 1, show the strength and direction of the relationship between two variables. A value of 1 indicates a perfect positive correlation, a value of -1 indicates a perfect negative correlation, and a value of 0 indicates no connection at all between the variables.

3. THE PROPOSED CNN MODEL FOR MONKEYPOX PREDICTION

In this stage, the CNN model is used to forecast monkeypox, motivated by the intriguing characteristics of deep networks. after the identification of the most prevalent characteristics and preprocessing of the monkeypox dataset. A 1D convolutional layer with a kernel size of two and the Rectified Linear Unit (ReLU) (Agarap, 2019) activation function makes up the input layer of the CNN architecture, which is depicted in Figure 3. The output layer is a dense layer with a single neurone and sigmoid activation function; the hidden layer is a dense layer with the ReLU activation function; the fatten layer is a layer to fatten the output from the preceding layer; and the pooling layer has a default pool size of two.

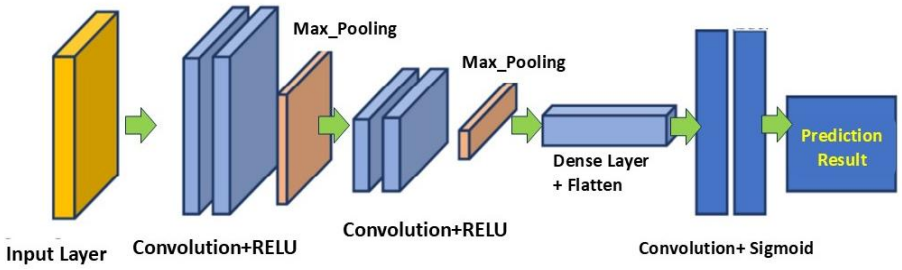

Figure 3: Architecture of the Proposed CNN Model

The particular CNN layers utilised in the architecture shown in Figure 3 were chosen based on the features of the dataset used in the experiment, and this study examined a number of CNN designs. Accuracy, loss, and validation results were some of the performance metrics used to optimise the various CNN designs. To make sure the system wasn't overfitted or underfitted, we used a range of strategies during the model's training and assessment phases. We ensured that the model was sufficiently sophisticated to identify the underlying patterns in the data in order to avoid underfitting. By choosing the right model architecture and hyperparameters, this was accomplished. We also pre-processed the data and added extra samples to the dataset to ensure that it was representative of the issue domain.

To prevent overfitting during the training phase, we used techniques like regularisation and early pausing. Early stopping was used to halt the training process as soon as the validation error stopped increasing, preventing the model from learning the training data. To reduce the model's complexity and prevent it from overfitting to the training set, regularisation was also used. We also evaluated the model's performance on an alternative test set to ensure that it worked well when applied to unknown data. If the model performed well on the test set, it indicated that it was not overfit to the training data. A trade-off between generalisation performance and model complexity was explored in order to avoid either overfitting or underfitting the system.

3.1 CNN Hyperparameter Optimization using GWO Algorithm

For our design, the CNN model was chosen, and its hyperparameters were changed to observe the outcomes. Since the deep learning architecture is composed of parameterised functions, the accuracy of monkeypox detection is greatly impacted by optimal parameters in particular. To find the optimal values for a range of hyperparameters, such as learning rate, batch size, number of layers, and filter size, the GWO technique (Mirjalili et al., 2014) has been utilised to address several optimisation problems, including CNN parameter tuning. The GWO optimisation approach may be used to modify CNN parameters by defining the search space for each hyperparameter and searching for the optimal set of hyperparameters (Mustaffa et al., 2015). The hyperparameters are seen as choice variables, while CNN's performance metric, classification accuracy, acts as the method's objective function (El-Gaafary et al., 2015). A population of grey wolves, each representing a potential solution, is the starting point for the GWO algorithm (Emary et al., 2015).

To ascertain the effectiveness of each solution, the CNN is trained on a training dataset and its performance is assessed on a validation dataset. A series of algorithms that mimic the social behaviour of grey wolves in the wild are then used to adjust each grey wolf's location based on how well each solution performs. This iterative process continues until a stopping condition is reached, such as a restriction on the number of iterations or a little improvement in the performance indicator. The optimal solution identified by the GWO approach is the same as the CNN's ideal set of hyperparameters. For monkeypox prediction, Figure 4 displays the flowchart for the proposed GWO-based hyperparameter-optimized CNN algorithm (Altindis et al., 2022; McCollum and Damon, 2014).

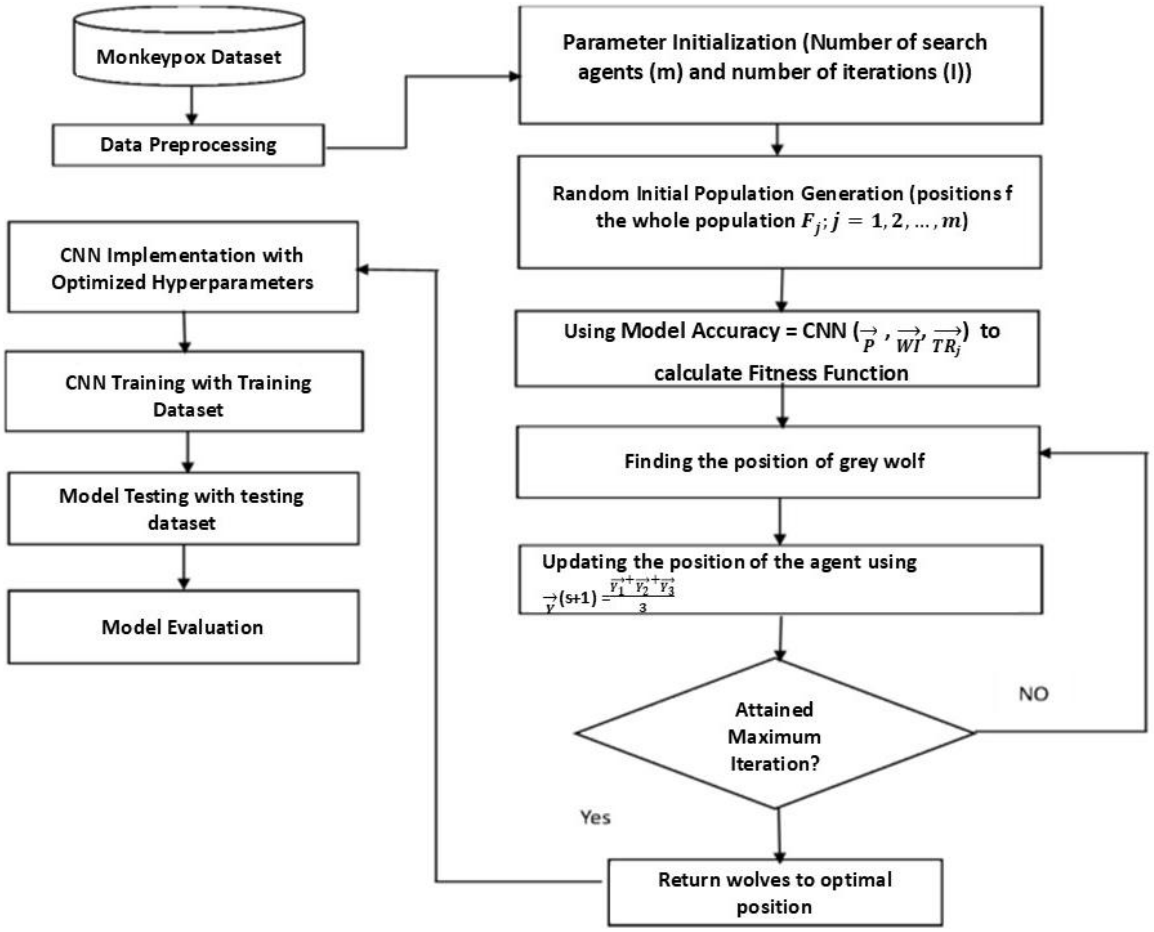

Figure 4: Flowchart of the proposed monkeypox prediction model

Figure 4 shows the CNN architecture, which accepts input vectors,

$\vec{P}$, $\vec{WI}$, $\vec{TR_j}$, where $\vec{P}$, represents the hyper-parameter vector of k dimension, $\vec{WI}$ stands for the weight vector of CNN, $\boldsymbol{TR_j}$ stands for some data that was selected from training data. The resultant output of this function is the accuracy of the model.

4. SYSTEM IMPLEMENTATION

A organised technique that includes CNN modelling, feature selection, data preparation, and hyperparameter optimisation using the Grey Wolf Optimisation (GWO) algorithm is used to develop the monkeypox prediction system in MATLAB. To improve model performance, data preparation entails loading the dataset, eliminating missing values, and balancing the dataset using the SMOTEEN approach. Correlation analysis is used in feature selection to find and keep the most important characteristics that are associated with the target variable, increasing accuracy and decreasing overfitting. An input layer, convolutional and pooling layers, fully connected layers and a sigmoid-activated output layer comprise the architecture of a CNN. To avoid underfitting and overfitting, the model is trained using the proper settings, such as regularisation and early stopping.

The GWO technique is used to optimise CNN hyperparameters such learning rate, batch size, and filter size in order to improve prediction accuracy. Using classification accuracy as the goal function, the GWO algorithm iteratively searches for the ideal parameter set by simulating the social behaviour of grey wolves. Following training on the processed and balanced dataset, the optimised CNN's performance is assessed on a different test set using measures like as accuracy, precision, recall, and F1-score. This guarantees that the model will generalise to new data, resulting in an effective and trustworthy approach for predicting monkeypox. For smooth system component execution and integration, the solution makes use of MATLAB's Deep Learning and Machine Learning Toolboxes.

5. RESULTS AND DISCUSSION

Prior to employing the CNN, the monkeypox dataset was pre-processed by eliminating null values and balancing the data using the SMOTEEN technique. Following the identification of the dataset's most prevalent characteristics, CNN hyperparameter optimisation was carried out using the GWO technique. to identify the crucial factors that affected the decision to diagnosis monkeypox. Figure 5 shows how the most prevalent variables in

the dataset were found using the proposed model: Rectal pain, fever, STIs, enlarged lymph nodes, sore throat, penile oedema, and oral lesions are all symptoms of HIV infection.

Figure 5: Correlation matrix of the Monkeypox diagnoses decision

After preprocessing the dataset and determining the most common features, the CNN model is utilised to predict monkeypox. However, the CNN architecture and hyperparameter settings have a significant influence on accuracy and convergence speed. Making these hyperparameter adjustments by hand is quite time-consuming and costly. Consequently, an automated approach is required to design the best CNN architecture with the least amount of human intervention. In this study, we used the GWO approach to optimise the CNN hyperparameters inside a predefined search area. To investigate the effect on the optimisation performance, we varied the maximum number of iterations and the number of search agents (population size). As indicated in Table 1, the optimised model was applied in three distinct scenarios to obtain the best result.

Table 1: Performance results of the CNN model optimized using the GWO algorithm with different population sizes

GWO with different parameter combinations	**Fitness score**	**Evaluation metrics for CNN model**				
		Accuracy	**Precision**	**Recall**	**F1 Score**	**AUC Score**
Population size = 50 Iteration = 10	0.93632	95.212	95.538	98.045	96.775	92.586
Population size=60 Iteration = 20	0.90252	93.365	93.706	97.622	95.624	89.156
Population size=70 Iteration = 30	0.90948	93.176	93.325	97.814	95.517	88.573

Table 1 displays the evaluation metrics for many CNN models with varied hyperparameters that were optimised using the GWO method. Various combinations of population size and iteration count were used to optimise these hyperparameters. Accuracy, precision, recall, F1 score, and AUC score were among the measures used to assess the models' performance.

The findings in Table 1 clearly show that the CNN model with the highest accuracy (95.212%) was trained with 10 iterations and a population size of 50 using parameters generated from GWO. This model also achieved the highest accuracy (95.538%) and recall (98.045%), indicating that it performed well in identifying positive cases and lowering false positives. The model also obtained a high AUC score of 92.586% and a high F1 score of 96.775%. The study's findings suggest that the hyperparameters used by GWO such as population size, learning rate, and number of iterations, which may have a substantial effect on CNN performance. As seen by the different fitness ratings attained in each run, it is also critical to keep in mind that the population

size and number of iterations employed in the optimisation process may have an impact on the final result. Interestingly, the study also shows that larger populations and more repeats did not always result in higher fitness scores.

Figure 6 shows the CNN model's ROC curve on the test dataset after its parameters were adjusted using the GWO approach. The GWO method was used with various population size and iteration count combinations. The ROC curve with an AUC score of 0.93 displays the CNN model that was trained with a population size of 50 and 10 iterations using the GWO optimiser. This result suggests that the model has a good discriminating potential and can effectively differentiate between positive and negative classes.

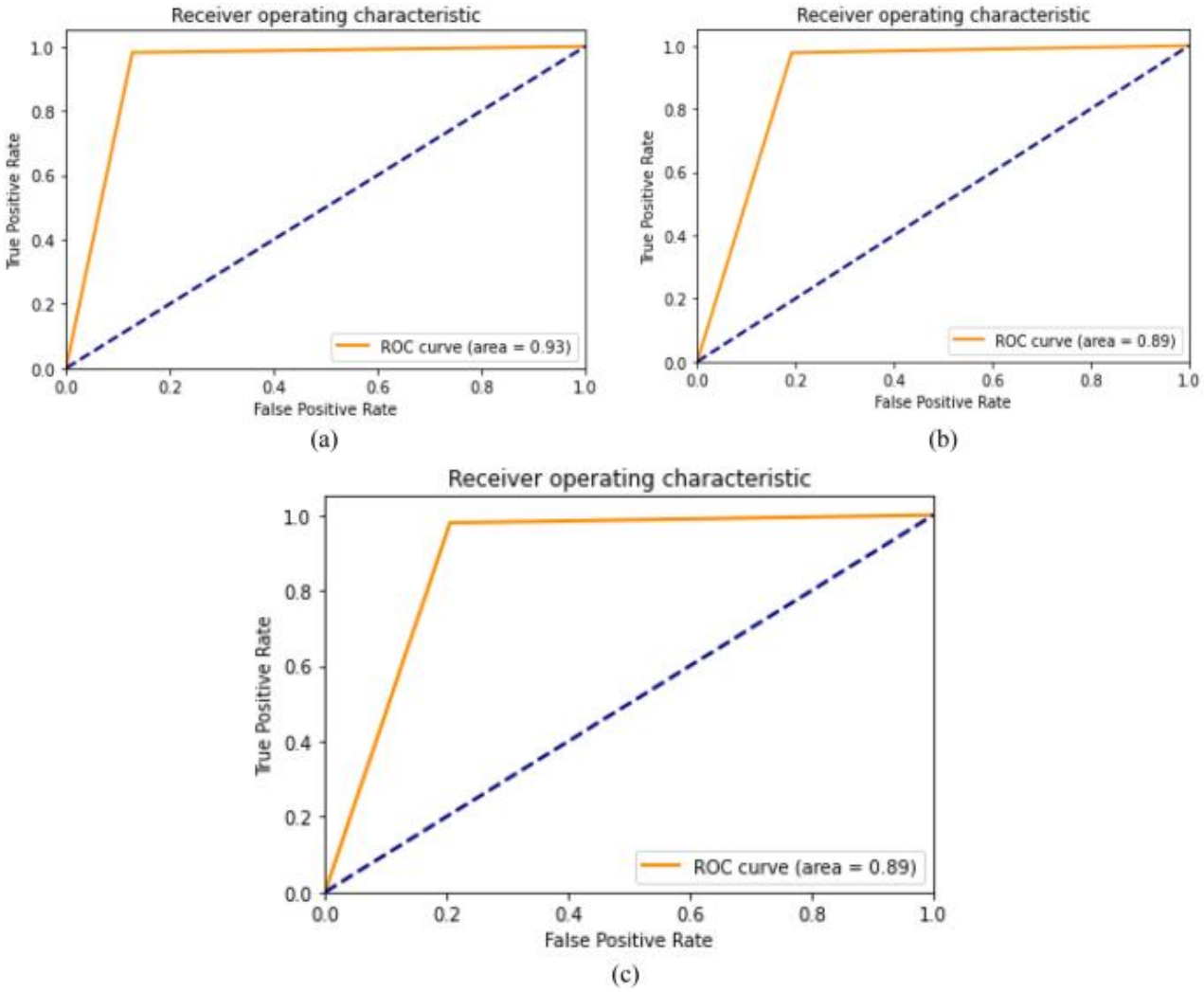

Figure 6: CNN model ROC curve after applying GWO in the three conditions (a) population size = 50 and iterations = 10 (b) population size = 60 and iterations = 20 (c) population size = 70 and iterations = 30

5. CONCLUSION

This study used a Convolutional Neural Network (CNN) optimised with the Grey Wolf Optimisation (GWO) method to create a reliable monkeypox prediction model. Preprocessing, feature selection, model creation, and hyperparameter tweaking were the main focusses of the technique. The dataset had 24,500 cases with 11 clinical characteristics and was obtained from Kaggle. The SMOTEEN technique, which reduced noise and addressed class imbalance, was used to balance the data and remove missing values as part of the preprocessing step. Eight crucial symptoms, including fever, swollen lymph nodes, and rectal pain, were shown to be the most suggestive of a monkeypox infection by feature selection using correlation analysis.

In order to avoid underfitting and overfitting, the CNN architecture was customised for the dataset using regularisation approaches, max pooling, and ReLU activation functions. In order to ensure optimal model performance, the GWO algorithm successfully adjusted important hyperparameters such as learning rate, batch size, and filter size. During GWO optimisation, a population size of 50 and 10 iterations produced the best results. With a 95.21% accuracy, 95.54% precision, 98.05% recall, and a 96.78% F1-score, the model showed exceptional performance. An Area Under the Curve (AUC) score of 92.59% demonstrated how well the model could differentiate between instances of monkeypox and those that weren't.

The study concludes that in order to create high-performing diagnostic systems, it is critical to combine sophisticated preprocessing methods, effective feature selection, and hyperparameter optimisation. The suggested method produced a dependable and accurate monkeypox prediction model by effectively addressing issues including class imbalance, noise, and parameter tuning. These results offer a foundation for improving infectious disease detection systems by utilising deep learning and optimisation methods.

REFERENCES

Agarap, A. F. (2019). Deep learning using rectified linear units (ReLU). *arXiv preprint* arXiv:1803.08375.

Alejo, R., Sotoca, J. M., Valdovinos, R. M., & Toribio, P. (2010). Edited nearest neighbor rule for improving neural networks classifications. In L. Zhang, B. L. Lu, & J. Kwok (Eds.), *Advances in Neural Networks – ISNN 2010* (Vol. 6063, pp. 329–338). Springer. https://doi.org/10.1007/978-3-642-13278-0_39

Altindis, M., Puca, E., & Shapo, L. (2022). Diagnosis of monkeypox virus—An overview. *Travel Medicine and Infectious Disease, 102459.* https://doi.org/10.1016/j.tmaid.2022.102459

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies, 11*(7), 1636.

El-Gaafary, A. A., Mohamed, Y. S., Hemeida, A. M., & Mohamed, A.-A. A. (2015). Grey wolf optimization for multi-input multi-output system. *Universal Journal of Communication and Network, 3*(1), 1–6.

Eliwa, E. H. I., El Koshiry, A. M., Abd El-Hafeez, T., & Farghaly, H. M. (2023). Utilizing convolutional neural networks to classify Monkeypox skin lesions. *Scientific Reports, 13,* 14495. https://doi.org/10.1038/s41598-023-41545-z

Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning, 113,* 4903–4923. https://doi.org/10.1007/s10994-022-06296-4

Emary, E., Zawbaa, H. M., Grosan, C., & Hassenian, A. E. (2014). Feature subset selection approach by gray-wolf optimization. In *Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 1–13). Springer.

Gupta A., Bhagat M., & Jain V., (2023). Blockchain-enabled healthcare monitoring system for early Monkeypox detection. The Journal of Supercomputing. 79. 1-25. 10.1007/s11227-023-05288-y.

Hahn, L. D., Baeumler, K., & Hsiao, A. (2021). Artificial intelligence and machine learning in aortic disease. *Current Opinion in Cardiology, 36*(6), 695–703.

Hussain, M. A., Islam, T., Chowdhury, F. U. H., & Islam, B. R. (2022). Can artificial intelligence detect Monkeypox from digital skin images? *bioRxiv.* https://doi.org/10.1101/2022.

Jaradat, A. S., Al Mamlook, R. E., Almakayeel, N., Alharbe, N., Almuflih, A. S., Nasayreh, A., Gharaibeh, H., Gharaibeh, M., Gharaibeh, A., & Bzizi, H. (2023). Automated Monkeypox skin lesion detection using deep learning and transfer learning techniques. *International Journal of Environmental Research and Public Health, 20*(4422). https://doi.org/10.3390/ijerph20054422

Kattenborn, T., Leitlöf, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 173,* 24–49.

Kawakami, E., Tabata, J., Yanaihara, N., Ishikawa, T., Koseki, K., Iida, Y., Saito, M., Komazaki, H., Shapiro, J. S., Goto, C., Akiyama, Y., Saito, R., Saito, M., Takano, H., Yamada, K., & Okamoto, A. (2019). Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. *Clinical Cancer Research, 25*(10), 3006–3015. https://doi.org/10.1158/1078-0432.CCR-18-3378

Keser, G., Bayrakdar, İ. Ş., Pekiner, F. N., Çelik, Ö., & Orhan, K. (2023). A deep learning algorithm for classification of oral lichen planus lesions from photographic images: A retrospective study. *Journal of Stomatology, Oral and Maxillofacial Surgery, 124*(1), 101264. https://doi.org/10.1016/j.jormas.2022.101264

Lin, D. J., Johnson, P. M., Knoll, F., & Lui, Y. W. (2021). Artificial intelligence for MR image reconstruction: An overview for clinicians. *Journal of Magnetic Resonance Imaging, 53*(4), 1015–1028.

McCollum, A. M., & Damon, I. K. (2014). Human monkeypox. *Clinical Infectious Diseases, 58*(2), 260–267. https://doi.org/10.1093/cid/cit703

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software, 69,* 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

Moore, M. J., Rathish, B., & Zahra, F. (2022). Mpox (Monkeypox). In *StatPearls*. StatPearls Publishing. Retrieved from https://www.statpearls.com

Muaz, A., Jayabalan, M., & Tiruchelvam, V. (2020). A comparison of data sampling techniques for credit card fraud detection. *International Journal of Advanced Computer Science and Applications, 11*(6).

Mustafa, Z., Sulaiman, M. H., & Kahar, M. N. M. (2015). LS-SVM hyper-parameters optimization based on GWO algorithm for time series forecasting. In *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)* (pp. 183–188). IEEE.

Thornhill, J. P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L. B., Palich, P., Nori, A., Reeves, I., Orkin, C. M., et al. (2022). Monkeypox virus infection in humans across 16 countries—April–June 2022. *New England Journal of Medicine, 387*(679–691). https://doi.org/10.1056/NEJMoa2207323

World Health Organization (WHO). (2023). Monkeypox. Available online: https://www.who.int/news-room/fact-sheets/detail/monkeypox (Accessed November 2024).