Machine Learning to Identify Monkey Pox Disease

Febri Aldi[1)*], Irohito Nozomi[2)], Rio Bayu Sentosa[3)], Ahmad Junaidi[4)]
[1,2,3,4)]Universitas Putra Indonesia YPTK Padang, Indonesia
[1)]febri_aldi@upiyptk.ac.id, [2)]irohito_nozomi@upiyptk.ac.id , [3)]riobayusentosa@upiyptk.ac.id,
[4)]ahmad_junaidi@upiyptk.ac.id

Abstract: In May 2022, it has received by WHO reports from non-endemic countries on cases of monkey pox disease. Monkey pox is a rare zoonotic disease caused by infection with the monkeypox virus that belongs to the genus orthopoxvirus and the family poxviridae, and also the variola virus. This study aims to classify patients who have contracted the monkey pox virus. We modeled an analysis of monkey pox disease and conducted comparisons utilizing a dataset from Kaggle consisting of a CSV file with records for 25,000 patients. The monkey pox dataset was analyzed using the correlation coefficient and the number of target variables. Machine learning (ML) methods are used for classification by utilizing the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB) algorithms. This study resulted in the highest classifier Gradient Boosting (GB) algorithm with an accuracy value of 71%. then the accuracy obtained by Support Vector Machine (SVM) is 69%, Random Forest (RF) accuracy is 68%, and finally K-Nearest Neighbor (KNN) obtains 63% accuracy. This ML method is expected to analyze monkey pox disease so that it helps the country and government, especially the health field in assessing, identifying, and being able to take appropriate action against monkey pox disease.

Keywords: Monkey Pox; Machine Learning; Kaggle; Classification; SVM

INTRODUCTION

In May 2022, it has received by WHO a report from non-endemic Countries on cases of monkey pox disease (Mohapatra et al., 2022). Currently, 12 non-endemic countries in three WHO regions namely Europe, America, and the Western Pacific are reported to have contracted monkey venom (Al Awaidy & Sallam, n.d.). The earliest cluster of cases was found in the UK, where the first case was found in a patient with a recent travel history from Nigeria on May 7, 2022, in London (Girometti et al., 2022). Monkeypox is a rare zoonotic disease caused by infection with monkeypox viruses belonging to the genus Orthopoxvirus and family Poxviridae, and also Variola virus, which causes smallpox, vaccinia virus, which is used in smallpox vaccines, and cowpox virus, all members of the genus orthopoxvirus (Besombes et al., 2019). The Democratic Republic of the Congo reported the first case of monkey pox in humans in 1970 (Durski et al., 2018). The symptoms of monkey pox in humans are comparable to the symptoms of chickenpox but not too severe. Fever, headache, muscle aches, and fatigue are the first signs . Identifying monkey pox disease can be done data analysis of the symptoms felt by patients using artificial intelligence (Kumar et al., 2022).

One of the artificial intelligences that can be used in identifying monkey pox disease is machine learning (Ahsan, Uddin, & Luna, 2022). Machine learning has a wide range of benefits in the healthcare industry, including disease identification, disease diagnosis, disease prediction, intelligent

*name of corresponding author

health records, medical imaging, and more (P. Singh et al., 2021). It is therefore understandable that many people believe that the existence of machine learning would be of great benefit. Many methods available in machine learning can be used to process data (Zitnik et al., 2019). These algorithms are grouped into three categories: reinforcement learning, supervised learning, and unsupervised learning, with each category subdivided into several types of algorithms (Saravanan & Sujatha, 2019). This study was conducted to classify monkey pox using supervised machine learning algorithms, which include KNN, SVM, RF, and GB. There have been a number of studies conducted on this case of monkey pox within the scope of machine learning methods, with key differences in objectives, time periods of disease collection, data types, and number of data sets.

LITERATURE REVIEW

As research conducted by Nenad Petrović (Petrovic & Petrović, n.d.) Examining the case of monkey pox in the Information Systems course using machine learning algorithms, that this study shows how machine learning techniques can be included in epidemic countermeasures information systems. The monkeypox case analysis dataset, which is available for free, is used for case studies, and the Java Weka API is used for implementation. Two predictive models of supervised learning were developed as a result and implemented into the Java Enterprise Edition architecture, showing satisfactory results in terms of prediction accuracy and execution speed.

Towhidul Islam et al (T. Islam, Hussain, Uddin, et al., 2022) conducted research using web scraping to create a thorough database of skin images affected by Measles, Cowpox, ChickenPox, Monkey Pox, and Smallpox. Their database has the most actual photos per class and the most enhanced images per class when compared to other comparable data sets. To verify the disease in the photo, they used the professional judgment of two doctors. In addition, they create lists to credit the creators of each image and make it available for use and inspection by the general public. They think that this database will make it easier to create basic ML and DL algorithms for early detection of Monkeypox, give ML or DL models a better chance of categorizing various skin lesions/rashes and aid overall clinical diagnosis of monkey pox.

Kolluri et al (Kolluri et al., 2022) conducted research by developing an open-source corpus of 225 cascading monkeypox claims using data from official World Health Organization (WHO) materials and fact-checking sources. Next, they developed an open-source BBERT based machine learning model to categorize information primarily related to monkeypox, which achieves a 96% cross-validation accuracy. As for Patel et al in the study stated that artificial intelligence in this case machine learning algorithms allows to be a useful tool in the future to improve patient care and can be a powerful weapon in the fight against monkey pox infection.

Towhidul Islam (T. Islam, Hussain, Chowdhury, et al., 2022) in its research mentions the scientific community has expressed growing interest in using artificial intelligence (AI) to diagnose monkey pox from digital skin photos as a result of AI's success in COVID-19 identification. The obstacle to using AI in monkeypox detection, however, is the scarcity of monkeypox skin image data. To test the feasibility of using cutting-edge AI deep models on skin photos for monkeypox detection, we used a web-scraping-based dataset consisting of images of healthy skin, chickenpox, smallpox, cowpox, measles, and monkey pox. According to our work, a deep AI model can identify monkeypox with an 85% accuracy rate of digital skin images. A larger training sample is needed to train those deep models to achieve stronger detection power.

Marwa Eid (Eid et al., 2022) in his research presented a new method for precise prediction of confirmed cases of monkey pox using tissue in optimized Long Short-Term Memory (LSTM). The proposed method is referred to as BER-LSTM because we use the Al-Biruni Earth Radius (BER) optimization algorithm to fine-tune the hyper parameters of LSTM-based deep networks. When the experimental findings were evaluated using other evaluation criteria, such as Mean Bias Error, which was measured with BER-LSTM and was (0.06), it was proved that the suggested approach was effective.

Research conducted by Abdelhamid (Abdelhamid et al., 2022) in image-based diagnostics, such as cancer detection, tumor cell identification, and COVID-19 patient detection, machine learning these

*name of corresponding author

days shows great hope. Therefore, when monkey pox spreads to human skin, similar techniques can be used to detect it. To help the subsequent diagnosis of the condition, images can be obtained. Two strategies were suggested in this study to improve the accuracy of monkeypox photo classification. The recommended approach is based on feature extraction with transfer learning, feature selection through meta-heuristic optimization, and parameter optimization for multi-layer neural networks. The findings obtained resulted in an average of 98.8% of classifications being made correctly.

Ahsan (Ahsan, Uddin, Farjana, et al., 2022) in its research in image-based diagnoses, such as cancer diagnosis, tumor cell identification, and detection of COVID-19 patients, machine learning (ML) has recently shown enormous potential. As a result, a similar application can be used to identify diseases associated with monkey pox as it affects human skin. These images can then be obtained and used to identify the disease. According to their first computational findings, the proposed model could accurately identify monkey pox patients in studies one and two with 97.18% (AUC=97.2) and 88.87% (AUC=0.867), respectively. In addition, they provide interpretable local model-agnostic (LIME) explanations for model prediction and feature extraction to provide a deeper understanding of the early typical characteristics of monkeypox viruses.

Researchers have recently suggested that machine learning is very beneficial in dealing with cases of monkey pox using different methods and types of data. Therefore, we want to take advantage of machine learning methods in classifying positive and negative monkey pox patient data that we obtained, using several classification algorithms that we will convey next. The aim is that medical personnel can take appropriate measures against monkey pox disease.

METHOD

The steps in the recommended framework for building an analytical model of monkey pox disease identification are presented in Figure 1. As shown in the figure, we begin the pre-processing step after downloading the dataset, which includes a number of data extracting, data cleansing, and data visualization. After pre-processing, we classified patients who were positive for monkey pox and negatively infected with monkey pox using a series of machine learning (ML) classification techniques. In terms of recall, precision, F-score, and accuracy, the process of performance assessment and comparison is then completed. The next subsection provides a thorough explanation of this action.
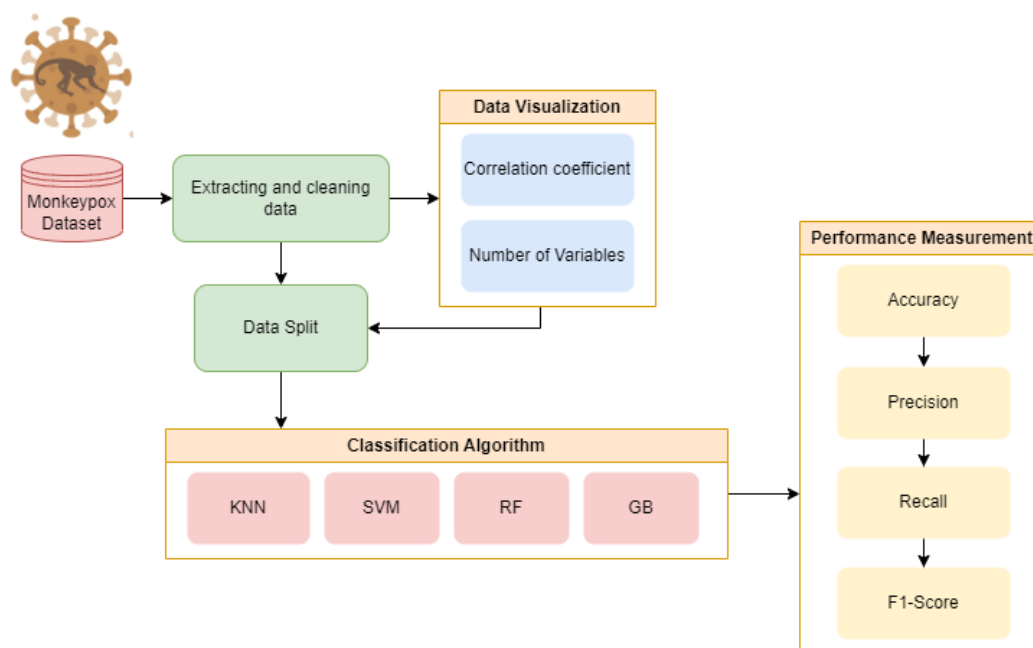

Fig. 1 Research methods

*name of corresponding author

Dataset

In this study, we modeled an analysis of monkey pox disease and made comparisons utilizing datasets from Kaggle (*Monkey-Pox PATIENTS Dataset. | Kaggle*, n.d.). The data collection consisted of a CSV file with records for 25,000 patients, each with their own attributes and target variables indicating whether or not they had monkey pox.

Extracting and Cleaning Data

Extracting and cleaning data in data mining is very important to do before the data is processed (Shah et al., 2020). The attribute values in the monkey pox data that are still letters will be converted into numbers so that the data can be processed properly.

Data Visualization

Use visual elements to display data, such as diagrams, graphs, or maps, known as data visualizations (Wang et al., 2020). Data visualization turns difficult-to-process numeric, complex, or high-volume data into visual representations (Divya Zion, 2020). The analytical intelligence of a disease is developed through data visualization, which also facilitates decision-making and data-driven strategy planning (Li et al., 2022). The two data visualization methods that we use for monkeypox patient data are the correlation coefficient and the number of variables. The correlation coefficient measures how strong the linear relationship between two variables is (Schober & Schwarte, 2018). The letter r is usually used to represent the correlation coefficient, which has a value between -1 and +1(Kim & Lakshmi, 2018). Knowing the number of variables in a target attribute is important to do (Chen et al., 2020).

Data Splitting

Data splitting is a technique for breaking down a set of data into two or more subsets (Roux, 2018). When the data is broken down into two parts, the first part is used to test or evaluate the data, while the second part is used to train the model (Vabalas et al., 2019). The purpose of building an ML model using training sets is to learn data patterns to generalize the model to new and unexplored data (Sarker, 2021). The second set is the test set, in which the model is equipped with fictitious data to forecast the results and check the output to assess the performance of the model (Ahmad et al., 2020).

Classification Algorithm

For the classification of monkey pox disease, we used a number of machine learning (ML) algorithms to conduct thorough studies, fine-tune performance, and compare the results. Our study used the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB) classifiers. We cover a quick summary of each classifier in this section.

K-Nearest Neighbor (KNN)

Easy and user-friendly machine learning, called K-Nearest Neighbor (KNN), can be used to address classification and regression issues (Shabani et al., 2020). The KNN algorithm belongs to the supervised learning category (J. U. Islam et al., 2019). The KNN algorithm is a machine learning technique that has non-parametric properties (Athani et al., 2021). The non-parametric approach suggests that it does not depend on any presumptions about how the underlying data will be distributed (Rogers et al., 2019). In other words, whether the data is small or large, the model does not have a fixed set of parameters or parameter estimates. The number of parameters used by non-parametric algorithms such as KNN can be adjusted and often increases as the amount of data grows (Saadatfar et al., 2020). Non-parametric algorithms take fewer assumptions about data but are computationally slower (Gautam & Singh, 2020). We use KNN to classify monkeypox and conduct experiments using the data we will evaluate.

*name of corresponding author

Support Vector Machine (SVM)

One of the techniques in supervised learning known as Support Vector Machine (SVM) is usually used for regression (Support Vector Regression) and classification (such as Support Vector Classification)(Cervantes et al., 2020). Compared to other classification techniques, SVM offers a more developed and mathematical idea for classification modeling (Horak et al., 2020). SVM can also solve linear and non-linear classification and regression problems (Zareef et al., 2020). We use SVM to classify monkeypox and conduct experiments using the data we will evaluate.

Random Forest (RF)

Large data sets can be classified using machine learning methods with Random Forest algorithms (Lakshmanaprabu et al., 2019). The RF algorithm uses decision trees to carry out a selection process, where decision trees will be recursively divided based on data of the same class (Herrera et al., 2019). More trees are used, will result in better accuracy. RF selects the best feature from a random subset of features while increasing model uncertainty (Sahin, 2020). This approach results in a large range and, generally, results in a better model. We use RF to classify monkeypox and conduct experiments using the data we will evaluate.

Gradient Boosting (GB)

A machine learning approach called gradient improvement leverages a group of decision trees to estimate value (U. Singh et al., 2021). Gradient boost is an optimization technique that uses boosting while utilizing the loss function to reduce errors (Neelakandan & Paulraj, 2020). We use GB to classify monkeypox and conduct experiments using the data we will evaluate.

Performance Measurement

Four different performance indicators are used to assess the performance of our model. By dividing the overall number of accurately predicted monkey pox by the total number of predicted monkey pox, accuracy is first determined. Equation 1 is used to calculate accuracy.

$$Akurasi = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (1)$$

Where the number of monkeypox cases that are projected to be positive and really positive is known as True Positives (TP). The number of monkey pox cases that are projected to be negative and actually negative is known as True Negative (TN). Monkeypox that is expected to be positive but turns out to be negative is referred to as a false positive (FP). Monkey pox that is projected to be negative but actually positive is referred to as a false negative (FN).

Second, the false positive classifier is calculated as a way to measure precision. Equation 2 is used to calculate precision.

$$Presisi = \frac{(TP)}{(TP + FP)} \qquad (2)$$

Third, the negative error of the classifier is calculated to determine the recall. Recall is determined by applying Equation 3.

$$Recall = \frac{(TP)}{(TP + FN)} \qquad (3)$$

The final stage, the weighted harmonic mean of the recall and precision is used to produce the F1 score. Equation 4 is used to determine the F1 score.

$$F1 - Score = \frac{2 \; x \; Precision \; x \; Recall}{Precision + Recall} \qquad (4)$$

*name of corresponding author

RESULT

The results of the monkeypox classification model outlined in the section above are presented in this section along with performance measurements. In addition to the four classifiers tested (KNN, SVM, RF, GB), we compared the performance of those models through precision accuracy, recall, and f1-score values. The results of our experiment are presented in the following sections:

Monkey Pox Dataset

The monkey pox dataset has attributes consisting of Patient ID, Systemic Disease, Rectal Pain, Sore Throat, Penile Edema, Oral Lesions, Solitary Lesions, Swollen Tonsils, HIV Infection, and Sexually Transmitted Infections, and the target variable is monkeypox. Table 1 is an example of the dataset used in this study.

Table 1
Monkey Pox Dataset

Patient ID	Systemic Diseases	Rectal Pain	sore throat	Penile edema	Oral Lesions	Solitary Lesions	swollen tonsils	HIV infection	Infectious Infections Seksual	Monkey Pox
P1	Not	False	True	True	True	False	True	False	False	Negative
P2	Fever	True	False	True	True	False	False	True	False	Positive
P3	Fever	False	True	True	False	False	False	True	False	Positive
P4	Not	True	False	False	False	True	True	True	False	Positive
P5	swollen lymph nodes	True	True	True	False	False	True	True	False	Positive
P6	swollen lymph nodes	False	True	False	False	False	False	False	False	Negative
P7	Fever	False	True	False	False	False	False	True	False	Positive
P8	Fever	True	True	False	True	True	True	False	False	Positive
P9	Muscle Aches and Pains	False	True	True	True	False	False	False	False	Positive
P10	Fever	False	False	True	True	True	False	True	False	Negative
: :	: :	: :	: :	: :	: :	: :	: :	: :	: :	: :
P25.000	swollen lymph nodes	False	False	True	False	False	True	True	False	Positive

This monkey pox data, which consisted of 25,000 patients, had systemic diseases such as fever, swollen lymph nodes, muscle pain, or had no systemic disease at all. Symptoms of a condition that affects one of the organs related to the metabolic system of the human body are called systemic diseases (Lu et al., 2019). On the attributes of rectal pain, sore throat, penile edema, oral lesions, solitary lesions, swollen tonsils, HIV infection, and sexually transmitted infections each have a true or false value. The monkey pox attribute as a target attribute has a positive or negative value.

*name of corresponding author

Extracting and Cleaning Data

The result of converting the value of each attribute in Table 2 shows the results of the conversions that have been made. In systemic disease attributes, the fever value is converted to the number 1.

Table 2
Monkey Pox Data Conversion Results

Patient ID	Syste mic Diseas es	Rect al Pain	sore throa t	Penil e edem a	Oral Lesion s	Solitar y Lesion s	swollen tonsils	HIV infecti on	Infectio us Infectio ns Seksual	Monk ey Pox
1	0	0	1	1	1	0	1	0	0	0
2	1	1	0	1	1	0	0	1	0	1
3	1	0	1	1	0	0	0	1	0	1
4	0	1	0	0	0	1	1	1	0	1
5	2	1	1	1	0	0	1	1	0	1
6	2	0	1	0	0	0	0	0	0	0
7	1	0	1	0	0	0	0	1	0	1
8	1	1	1	0	1	1	1	0	0	1
9	3	0	1	1	1	0	0	0	0	1
10	1	0	0	1	1	1	0	1	0	0
:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:
25.000	2	0	0	1	0	0	1	1	0	1

The value of swollen lymph nodes is converted to the number 2. The muscle pain value is converted to the number 3, and the value of no systemic disease is converted to the number 0. As for the attributes of rectal pain, sore throat, penile edema, oral lesions, solitary lesions, swollen tonsils, HIV infection, and sexually transmitted infections, which have a true value converted to the number 1, and false converted to the number 0. On the positive value on the monkey pox attribute is converted to the number 1, and the negative value is converted to the number 0. The accuracy of the model is directly affected by data cleansing and the formation of data analysis of smallpox monkey (Wang et al., 2020). We removed the Patient ID attribute as our step in cleaning up unnecessary data

Data Visualization
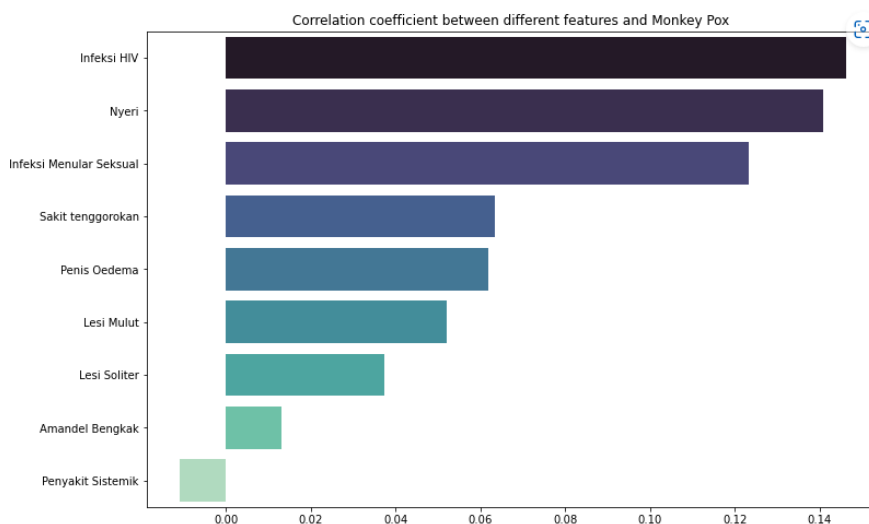Correlation Coefficient

Fig. 2 correlation coeffi

*name of corresponding author

Looking at the correlation between these variables shows a very weak relationship. Because, if the correlation value is between 0 – 0.25, it will result in a very weak correlation. On systemic disease variables have no correlation, since they are worth 0. The highest correlation in HIV infection variables with a value of 0 – 0.14.

Number of Variables

This aims to see the difference in the composition of an attribute so that it can be analyzed properly. Figure 3 is the number of variables on the monkey pox attribute.
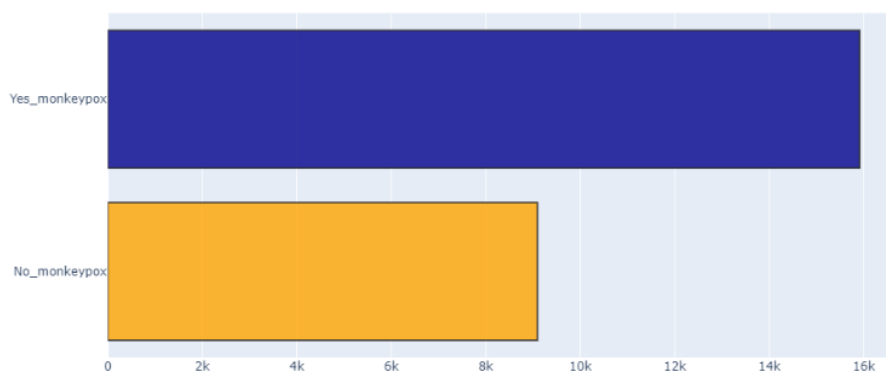

Fig. 3 Variable Number of Monkey Pox

A considerable difference is seen in Figure 3, where the number of variables in monkey pox positive patients outperforms those in monkey pox negative patients. Of the 25,000 patients, 15,909 patients were positive for monkey pox, and 9091 patients were negative for monkey pox.

Data Splitting

We split that data into 80% for training data, and 20% for test data. Such data separation applies to all algorithms to be tested.

Performance Measurement

	precision	recall	f1-score	support
0	0.64	0.37	0.47	2262
1	0.71	0.88	0.79	3988
accuracy			0.70	6250
macro avg	0.68	0.63	0.63	6250
weighted avg	0.69	0.70	0.67	6250

Fig. 4 Gradient Boosting classification results

Figure 4 shows the classification results of the Gradient Boosting algorithm with a precision of 71% positive, and 64% negative. On recall 88% were positive, and 37% were negative. On F1-Score 78% positive, and 47% negative. The accuracy value obtained is 70%.

	precision	recall	f1-score	support
0	0.48	0.43	0.45	2262
1	0.69	0.74	0.72	3988
accuracy			0.63	6250
macro avg	0.59	0.58	0.58	6250
weighted avg	0.62	0.63	0.62	6250

Fig. 5 KNN classification results

*name of corresponding author

Figure 5 shows the results of the KNN algorithm classification with a precision of 69% positive, and 48% negative. On recall 74% were positive, and 43% were negative. On F1-Score 72% positive, and 45% negative. The accuracy value obtained is 63%.

```
              precision   recall  f1-score   support

          0       0.59      0.37      0.46      2262
          1       0.70      0.85      0.77      3988

   accuracy                           0.68      6250
  macro avg       0.65      0.61      0.61      6250
weighted avg      0.66      0.68      0.66      6250
```

Fig. 6 Random Forest classification results

Figure 6 shows the classification results of the Random Forest algorithm with a precision of 70% positive, and 59% negative. On recall 85% were positive, and 37% were negative. On F1-Score 77% positive, and 46% negative. The accuracy value obtained is 68%.

```
              precision   recall  f1-score   support

          0       0.67      0.30      0.42      2262
          1       0.70      0.91      0.79      3988

   accuracy                           0.69      6250
  macro avg       0.68      0.61      0.60      6250
weighted avg      0.69      0.69      0.66      6250
```

Fig. 7 SVM classification results

Figure 7 shows the results of SVM algorithm classification with precision 70% positive, and 67% negative. On recall 91% were positive, and 30% were negative. On F1-Score 79% positive, and 42% negative. The accuracy value obtained is 69%.

DISCUSSION

After we measure the performance of the four classifiers above. Table 3 displays the performance results of the classification algorithm in terms of accuracy, precision, recall, and f1-score. The GB classifier outperformed the KNN, SVM, and RF classifiers, as it achieved the highest score: 70% accuracy.

Table 2.
Results of Four Classification Models

Classifiers	Accuracy	Precision		Recall		F1-Score	
		Positif	Negatif	Positif	Negatif	Positif	Negatif
KNN	63%	69%	48%	74%	43%	72%	45%
SVM	69%	70%	67%	91%	30%	79%	42%
RF	68%	70%	59%	85%	37%	77%	46%
GB	70%	71%	64%	88%	37%	79%	47%

*name of corresponding author

The highest monkey pox positive precision result lies in the GB classifier with a value of 71%, outperforming SVM and RF 70% as the second, and third KNN 69%. While the highest monkey pox negative precision of 67% was achieved by SVM, second GB 64%, third RF 59%, and fourth KNN 48%. Then the highest positive monkey pox recall results are located at SVM with a value of 91%, second GB 88%, third RF 85%, and fourth KNN 74%. While the highest monkey pox negative recall of 6743% was achieved by KNN, both RF and GB 37%, and third SVM 30%. The highest positive monkey pox f1-score on SVM and GB classifier with a value of 79%, second RF 77%, and third 72%. Then the highest negative monkey pox F1-Score at GB 47%, second RF 46%, third KNN 45%, and fourth SVM 42%. Looking at the overall results that we have tested on the ML model using four monkeypox classification algorithms when viewed from the f1-score results, that the GB and SVM algorithms are ranked first, RF ranks second, and KNN ranks third. Therefore, the overall results show not too bad results.

CONCLUSION

The purpose of this study was to test several ML models and evaluate their efficacy in analyzing monkey pox disease. Testing is carried out by testing various models using as a classifying tool. We investigate the performance of KNN, SVM, RF, and GB algorithms through performance measurements that display accuracy, precision, recall, f1-score, and time taken values. Given that the highest accuracy was 70%, the study showed that GB beat KKN, SVM, and RF in the analysis of monkey pox disease.

The methodology of monkey pox disease analysis can help countries and governments, especially the health field in assessing the impact of monkey pox disease and preventive measures on the health of patients and communities so that they can act quickly to deal with this monkey pox disease. In addition, both the health service and the health sector can use the model to anticipate and take swift action in addressing the monkey pox problem before it worsens and impacts the health of the world's people.

In the future, we intend to apply our model to the development of monkey pox disease. We can also investigate other useful research avenues for the analysis of this monkey pox disease. We intend to analyze the performance of deep learning classifiers (DL), then compare them to machine learning (ML) classifiers, in order to get better results.

REFERENCES

Abdelhamid, A. A., El-Kenawy, E. S. M., Khodadadi, N., Mirjalili, S., Khafaga, D. S., Alharbi, A. H., Ibrahim, A., Eid, M. M., & Saber, M. (2022). Classification of Monkeypox Images Based on Transfer Learning and the Al-Biruni Earth Radius Optimization Algorithm. *Mathematics 2022, Vol. 10, Page 3614*, *10*(19), 3614. https://doi.org/10.3390/MATH10193614

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. *Complexity*, *2020*. https://doi.org/10.1155/2020/8885861

Ahsan, M. M., Uddin, M. R., Farjana, M., Sakib, A. N., Momin, K. Al, & Luna, S. A. (2022). *Image Data collection and implementation of deep learning-based model in detecting Monkeypox disease using modified VGG16*. https://doi.org/10.48550/arxiv.2206.01862

Ahsan, M. M., Uddin, M. R., & Luna, S. A. (2022). *Monkeypox Image Data collection*. https://doi.org/10.48550/arxiv.2206.01774

Al Awaidy, S., & Sallam, M. (n.d.). *Systematic Review and Meta-analysis on COVID-19 Vaccine Hesitancy View project Respiratory Viruses View project*. https://doi.org/10.18295/squmj.8.2022.046

Athani, S., Joshi, S., Rao, B. A., Rai, S., & Kini, N. G. (2021). Parallel Implementation of kNN Algorithm for Breast Cancer Detection. *Advances in Intelligent Systems and Computing*, *1176*, 475–483. https://doi.org/10.1007/978-981-15-5788-0_46/COVER

Besombes, C., Gonofio, E., Konamna, X., Selekon, B., Gessain, A., Berthet, N., Manuguerra, J. C., Fontanet, A., & Nakouné, E. (2019). Intrafamily Transmission of Monkeypox Virus, Central

*name of corresponding author

African Republic, 2018. *Emerging Infectious Diseases*, *25*(8), 1602. https://doi.org/10.3201/EID2508.190112

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/J.NEUCOM.2019.10.118

Chen, R. C., Dewi, C., Huang, S. W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1), 1–26. https://doi.org/10.1186/S40537-020-00327-4/FIGURES/13

Divya Zion, G. (2020). Comparative analysis of tools for big data visualization and challenges. *Data Visualization: Trends and Challenges Toward Multidisciplinary Perception*, 33–52. https://doi.org/10.1007/978-981-15-2282-6_3/COVER

Durski, K. N., McCollum, A. M., Nakazawa, Y., Petersen, B. W., Reynolds, M. G., Briand, S., Djingarey, M. H., Olson, V., Damon, I. K., & Khalakdina, A. (2018). Emergence of Monkeypox — West and Central Africa, 1970–2017. *Morbidity and Mortality Weekly Report*, *67*(10), 306. https://doi.org/10.15585/MMWR.MM6710A5

Eid, M. M., El-Kenawy, E. S. M., Khodadadi, N., Mirjalili, S., Khodadadi, E., Abotaleb, M., Alharbi, A. H., Abdelhamid, A. A., Ibrahim, A., Amer, G. M., Kadi, A., & Khafaga, D. S. (2022). Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox Cases. *Mathematics 2022, Vol. 10, Page 3845*, *10*(20), 3845. https://doi.org/10.3390/MATH10203845

Gautam, A., & Singh, V. (2020). Parametric Versus Non-Parametric Time Series Forecasting Methods: A Review. *Journal of Engineering Science and Technology Review*, *13*(3), 165–171. https://doi.org/10.25103/jestr.133.18

Girometti, N., Byrne, R., Bracchi, M., Heskin, J., McOwan, A., Tittle, V., Gedela, K., Scott, C., Patel, S., Gohil, J., Nugent, D., Suchak, T., Dickinson, M., Feeney, M., Mora-Peris, B., Stegmann, K., Plaha, K., Davies, G., Moore, L. S. P., … Whitlock, G. (2022). Demographic and clinical characteristics of confirmed human monkeypox virus cases in individuals attending a sexual health centre in London, UK: an observational analysis. *The Lancet Infectious Diseases*, *22*(9), 1321–1328. https://doi.org/10.1016/S1473-3099(22)00411-X

Herrera, V. M., Khoshgoftaar, T. M., Villanustre, F., & Furht, B. (2019). Random forest implementation and optimization for Big Data analytics on LexisNexis's high performance computing cluster platform. *Journal of Big Data*, *6*(1), 1–36. https://doi.org/10.1186/S40537-019-0232-1/FIGURES/17

Horak, J., Vrbka, J., & Suler, P. (2020). Support Vector Machine Methods and Artificial Neural Networks Used for the Development of Bankruptcy Prediction Models and their Comparison. *Journal of Risk and Financial Management 2020, Vol. 13, Page 60*, *13*(3), 60. https://doi.org/10.3390/JRFM13030060

Islam, J. U., Hollebeek, L. D., Rahman, Z., Khan, I., & Rasool, A. (2019). Customer engagement in the service context: An empirical investigation of the construct, its antecedents and consequences. *Journal of Retailing and Consumer Services*, *50*(May), 277–285. https://doi.org/10.1016/j.jretconser.2019.05.018

Islam, T., Hussain, M. A., Chowdhury, F. U. H., & Islam, B. M. R. (2022). Can Artificial Intelligence Detect Monkeypox from Digital Skin Images? *BioRxiv*, 2022.08.08.503193. https://doi.org/10.1101/2022.08.08.503193

Islam, T., Hussain, M. A., Uddin, F., Chowdhury, H., & Islam, B. M. R. (2022). A Web-scraped Skin Image Database of Monkeypox, Chickenpox, Smallpox, Cowpox, and Measles. *BioRxiv*, 2022.08.01.502199. https://doi.org/10.1101/2022.08.01.502199

Kim, H., & Lakshmi, V. (2018). Use of Cyclone Global Navigation Satellite System (CyGNSS) Observations for Estimation of Soil Moisture. *Geophysical Research Letters*, *45*(16), 8272–8282. https://doi.org/10.1029/2018GL078923

Kolluri, A., Vinton, K., & Murthy, D. (2022). *PoxVerifi: An Information Verification System to Combat Monkeypox Misinformation*. https://doi.org/10.48550/arxiv.2209.09300

*name of corresponding author

Kumar, N., Acharya, A., Gendelman, H. E., & Byrareddy, S. N. (2022). The 2022 outbreak and the pathobiology of the monkeypox virus. *Journal of Autoimmunity*, *131*, 102855. https://doi.org/10.1016/J.JAUT.2022.102855

Lakshmanaprabu, S. K., Shankar, K., Ilayaraja, M., Nasir, A. W., Vijayakumar, V., & Chilamkurti, N. (2019). Random forest for big data classification in the internet of things using optimal features. *International Journal of Machine Learning and Cybernetics 2019 10:10*, *10*(10), 2609–2618. https://doi.org/10.1007/S13042-018-00916-Z

Li, C., Chen, Y., & Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal*, *29*, 101021. https://doi.org/10.1016/J.JESTCH.2021.06.001

Lu, M., Xuan, S., & Wang, Z. (2019). Oral microbiota: A new view of body health. *Food Science and Human Wellness*, *8*(1), 8–15. https://doi.org/10.1016/J.FSHW.2018.12.001

Mohapatra, R. K., Tuli, H. S., Sarangi, A. K., Chakraborty, S., Chandran, D., Chakraborty, C., & Dhama, K. (2022). Unexpected sudden rise of human monkeypox cases in multiple non-endemic countries amid COVID-19 pandemic and salient counteracting strategies: Another potential global threat? *International Journal of Surgery (London, England)*, *103*, 106705. https://doi.org/10.1016/J.IJSU.2022.106705

Monkey-Pox PATIENTS Dataset. | Kaggle. (n.d.). Retrieved November 17, 2022, from https://www.kaggle.com/datasets/muhammad4hmed/monkeypox-patients-dataset

Neelakandan, S., & Paulraj, D. (2020). A gradient boosted decision tree-based sentiment classification of twitter data. *Https://Doi.Org/10.1142/S0219691320500277*, *18*(4). https://doi.org/10.1142/S0219691320500277

Petrovic, N., & Petrović, N. (n.d.). *Machine Learning within Information Systems Course using Weka in Java: Monkeypox Case Studies SCOR (Semantic COordination for Rawfie) View project Machine Learning within Information Systems Course using Weka in Java: Monkeypox Case Studies*. https://doi.org/10.13140/RG.2.2.26822.75842

Rogers, T. J., Worden, K., Fuentes, R., Dervilis, N., Tygesen, U. T., & Cross, E. J. (2019). A Bayesian non-parametric clustering approach for semi-supervised Structural Health Monitoring. *Mechanical Systems and Signal Processing*, *119*, 100–119. https://doi.org/10.1016/J.YMSSP.2018.09.013

Roux, M. (2018). A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms. *Journal of Classification 2018 35:2*, *35*(2), 345–366. https://doi.org/10.1007/S00357-018-9259-9

Saadatfar, H., Khosravi, S., Joloudari, J. H., Mosavi, A., & Shamshirband, S. (2020). A New K-Nearest Neighbors Classifier for Big Data Based on Efficient Data Pruning. *Mathematics 2020, Vol. 8, Page 286*, *8*(2), 286. https://doi.org/10.3390/MATH8020286

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, *2*(7), 1–17. https://doi.org/10.1007/S42452-020-3060-1/TABLES/4

Saravanan, R., & Sujatha, P. (2019). A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018*, 945–949. https://doi.org/10.1109/ICCONS.2018.8663155

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 1–21. https://doi.org/10.1007/S42979-021-00592-X/FIGURES/11

Schober, P., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T., & Várkonyi-Kóczy, A. R. (2020). Modeling Pan Evaporation Using Gaussian Process Regression K-Nearest Neighbors Random Forest and Support Vector Machines; Comparative Analysis. *Atmosphere 2020, Vol. 11, Page 66*, *11*(1), 66. https://doi.org/10.3390/ATMOS11010066

*name of corresponding author

Shah, D., Patel, S., & Bharti, S. K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science 2020 1:6*, *1*(6), 1–6. https://doi.org/10.1007/S42979-020-00365-Y

Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. *Machine Learning and the Internet of Medical Things in Healthcare*, 89–111. https://doi.org/10.1016/B978-0-12-821229-5.00003-3

Singh, U., Rizwan, M., Alaraj, M., & Alsaidan, I. (2021). A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments. *Energies 2021, Vol. 14, Page 5196*, *14*(16), 5196. https://doi.org/10.3390/EN14165196

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), e0224365. https://doi.org/10.1371/JOURNAL.PONE.0224365

Wang, T., Ke, H., Zheng, X., Wang, K., Sangaiah, A. K., & Liu, A. (2020). Big Data Cleaning Based on Mobile Edge Computing in Industrial Sensor-Cloud. *IEEE Transactions on Industrial Informatics*, *16*(2), 1321–1329. https://doi.org/10.1109/TII.2019.2938861

Zareef, M., Chen, Q., Hassan, M. M., Arslan, M., Hashim, M. M., Ahmad, W., Kutsanedzie, F. Y. H., & Agyekum, A. A. (2020). An Overview on the Applications of Typical Non-linear Algorithms Coupled With NIR Spectroscopy in Food Analysis. *Food Engineering Reviews 2020 12:2*, *12*(2), 173–190. https://doi.org/10.1007/S12393-020-09210-7

Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, *50*, 71–91. https://doi.org/10.1016/J.INFFUS.2018.09.012

*name of corresponding author