Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13967
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison of Machine Learning Algorithms on Stunting Detection for 'Centing' Mobile Application to Prevent Stunting

Ferris Tita Sabilillah[1)*], Christy Atika Sari[2)], Ryandhika Bintang Abiyyi[3)], Yani Parti Astuti[4)]
[1,2,3,4)] Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
[1)]111202113579@mhs.dinus.ac.id, [2)]christy.atika.sari@dsn.dinus.ac.id, [3)]111202113498@mhs.dinus.ac.id, [4)]yanipartiastuti@dsn.dinus.ac.id

Submitted :Aug 2, 2024 | **Accepted** : Aug 21, 2024 | **Published** : Oct 7, 2024

Abstract: Stunting is a growth disorder caused by chronic undernutrition, with long-term impacts on child health and development. In Indonesia, the prevalence of stunting reached 31.8% in children under five years old in 2018, indicating an urgent need for effective interventions. In an effort to address this issue, we developed a mobile application called Centing (Cegah Stunting) that utilizes machine learning for early detection and prevention of stunting. In this study, we compared the performance of four machine learning algorithms, namely Logistic Regression, Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, Convolutional Neural Network (CNN), and Multilayer Perceptron (MLP), in detecting children's nutritional status based on a dataset from Kaggle consisting of 121 thousand data with four main features, namely age, gender, height, and nutritional status. The experimental results show that SVM with RBF kernel and CNN achieved the highest accuracy of 98%, while Logistic Regression and MLP achieved 76% and 97% accuracy, respectively. SVM with RBF kernel was chosen as the best model due to its high accuracy and computational time efficiency. The selected SVM model with RBF kernel was then converted into TensorFlow Lite (TFLite) format to be efficiently integrated into the Centing mobile application, so that it can be optimally run on resource-constrained devices, such as smartphones. The findings show that the Centing app, with the implementation of SVM RBF, has significant potential in early detection and prevention of stunting, and makes an important contribution to improving child health in Indonesia.

Keywords: Stunting, Machine Learning, SVM, CNN, MLP, Logistic Regression

INTRODUCTION

The high prevalence of stunting has a long-term impact on children's physical and cognitive development. According to Asian Development Bank data, the prevalence of stunting in Indonesia in 2018 reached around 30.8% of the total children under the age of five. Based on the results of the Indonesian Nutrition Status Study (SSGI) in 2019, the prevalence of stunting in Indonesia is still above 20%, namely 24.4%, underweight 17%, and wasted 7.1% (Pitayanti et al., 2022). These figures indicate that almost one in three children in Indonesia is stunted, placing Indonesia high in the global stunting rankings and making it the third country with the highest cases in Asia according to the World Health Organization (WHO) (Sugianto, 2021). In 2019, the Indonesian Ministry of Health will concentrate integrated stunting intervention efforts in 160 districts across Indonesia (Wulandari & Kusumastuti, 2020). The problem of stunting has a significant impact on the quality of human resources, both in the short and long term (Rahman et al., 2023). Stunting threatens Indonesia's human quality and competitiveness as it causes impaired physical growth and brain development. This greatly affects ability and achievement at school, productivity, and creativity at productive age (Aurima et al., 2021). Stunting in children can inhibit brain development, resulting in suboptimal intelligence and educational achievement in adulthood and increasing the risk of metabolic diseases (Hamzah & B, 2020). Toddlerhood is a period that is very sensitive to the environment, so more attention is needed, especially to nutritional adequacy (Zurhayati & Hidayah, 2022). Tackling stunting is critical to improving the quality of life and future of future generations, yet major challenges still exist in Indonesia, such as limited access to health services, lack of nutrition knowledge, and difficulties in monitoring child growth. Therefore, innovative solutions are needed. Information technology and artificial intelligence offer more efficient and accurate solutions. This research aims to develop a mobile application called Centing (Cegah Stunting) that uses machine learning to detect and provide recommendations related to stunting in children in Indonesia. Various

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13967
e-ISSN : 2541-2019
p-ISSN : 2541-044X

machine learning algorithms, such as Logistic Regression, Support Vector Machine (SVM) with various kernels, Convolutional Neural Network (CNN), and Multi-Layer Perceptron (MLP), were evaluated to determine the most accurate and efficient model. Stunting has long-term impacts on quality of life, including cognitive delays, reduced productivity, and increased risk of chronic diseases in adulthood. Health services struggle with data access, child monitoring, and timely intervention. Machine learning is effective in stunting detection with the ability to process large data quickly and accurately. The Centing (Cegah Stunting) application utilizes this technology to provide more targeted recommendations, facilitate real-time monitoring of children's nutritional status, and offer advantages over traditional methods that tend to be slow and less accurate. With the application of machine learning models, Centing can detect the risk of stunting earlier and provide the necessary interventions, thus significantly contributing to efforts to improve child health in Indonesia.

LITERATURE REVIEW

Stunting

Nutritional status is a condition of balance between nutrient intake and nutrient requirements needed by the body for growth and development, especially for children under five. It also includes activity, health maintenance, healing for the sick, and other biological processes in the body (Wahyuni & Fithriyana, 2020). Stunting is a condition of chronic malnutrition in children that causes physical growth and brain development to be inhibited, so that their height is lower than their age standard. Stunting is measured by a height-for-age (TB/U) z-score of less than -2 standard deviations (SD) based on WHO growth standards. Children under five are at an age that is vulnerable to various diseases and nutritional problems (Nugroho et al., 2021). Children who are severely stunted or very short have a height-for-age z-score of less than -3 standard deviations (SD), while children are considered normal if the height-for-age (TB/U) z-score is more than -2 standard deviations (SD) (Zulfikar Lating et al., 2023). Stunting in Indonesia is caused by poor nutrition, infections and poor parenting practices, impacting cognitive development, productivity and increasing chronic diseases. Economically, stunting reduces potential earnings, slows growth, and exacerbates social inequality. Addressing it is essential for sustainable development.

Machine Learning

With the problem of stunting, a breakthrough is needed to deal with the factors that cause it quickly and accurately. One approach that can be used is to utilize the help of machine learning (Banurea et al., 2023). Machine learning is a branch of artificial intelligence that allows computers to learn from data and make predictions or decisions without having to be explicitly programmed. There are many machine learning methods that can be used to develop effective prediction models (Fahri & Ramdhani, 2023). Machine learning can efficiently identify and address stunting factors through data patterns. Applications in healthcare include diagnosis, medical image analysis, and personalized care. For example, algorithms such as Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) have been used in health data analysis to detect certain diseases and conditions with high accuracy. However, the use of machine learning technology must still be supported by data and recommendations from health professionals (Hakim et al., 2023).

Logistic Regression

Logistic regression is one of the most well-known machine learning algorithms after linear regression. While both have similarities, the main difference lies in their usage. Linear regression is used to predict values, while logistic regression is used for classification tasks. Logistic regression models the probability of a particular event as a function of one or more independent variables, transforming those probabilities into binary or multiclass categories (Gunawan, Muhammad Ichsan Sugiarto & Mardianto, 2020). The downside of Logistic Regression is its limited ability to handle highly complex and non-linear data. Logistic Regression has major advantages in terms of interpretability and computational speed. The model can be easily interpreted because the output is a probability that can be converted into binary or multiclass categories. The computational speed of Logistic Regression is also relatively fast, making it an efficient choice for applications with limited processing time. However, the drawback of Logistic Regression lies in its limited ability to handle highly complex and non-linear data.

SVM

SVM is a machine learning algorithm for classification and regression that searches for the optimal hyperplane to separate data into different classes with maximum margin, often using kernel functions such as RBF for data that cannot be linearly separated. Support Vector Machine (SVM) is a powerful algorithm for classification, especially with non-linear kernels. Support Vector Machine (SVM) algorithm, which has been proven accurate in health data analysis (Sanhaji et al., 2024).

CNN

One of the proposed solutions is the utilization of Convolutional Neural Networks (CNN) in Android applications (Sayyidin et al., 2024). CNN is an artificial neural network used for image and video recognition,

* Corresponding author

consisting of a convolution layer that extracts local features from the input, a pooling layer that reduces the dimensionality, and a fully connected layer that performs the final classification.

MLP

A Multilayer Perceptron (MLP) is an artificial neural network consisting of multiple layers of neurons, including an input layer, one or more hidden layers, and an output layer, all of which are fully connected. The MLP training process uses the backpropagation method to adjust weights based on prediction error, and is used for classification and regression tasks. In the MLP structure, the data flow starts from the input layer until it reaches the output layer without any repetition (Widodo et al., 2020).

METHOD

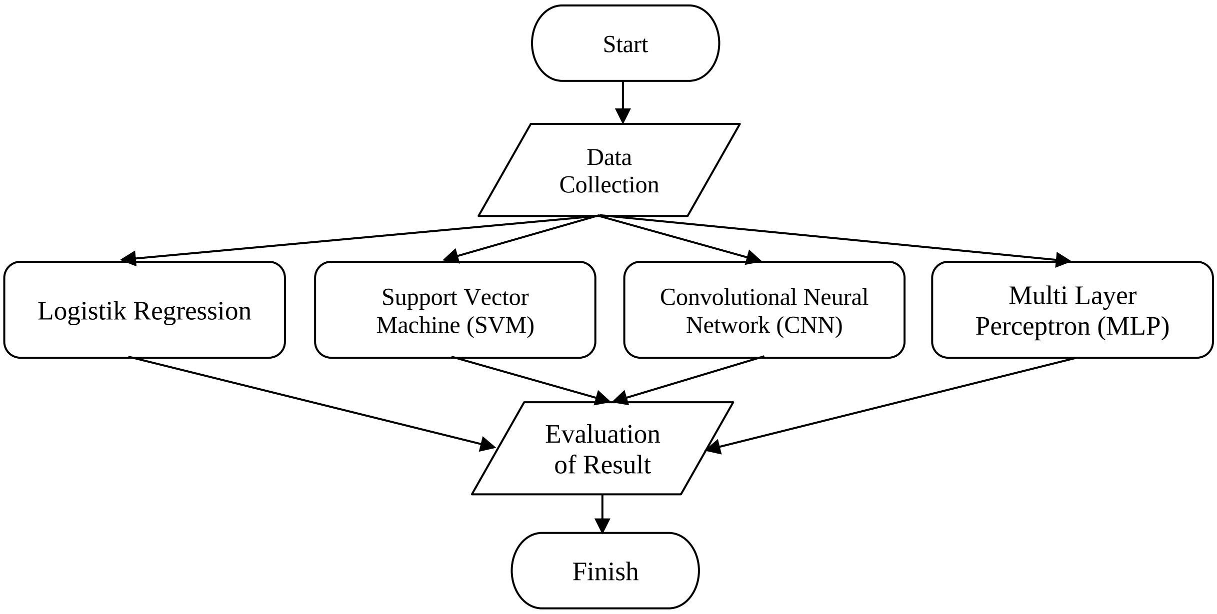

Fig. 1 Research flow. Created by Author, 2024

Data Collection

The data used in this study was obtained from a dataset available on Kaggle. The dataset consists of 121 rows of data that include features such as age (months), gender, height (cm), and nutritional status classified into four categories: severely stunted, stunted, normal (good growth), and tall (excellent growth).

Data Pre-processing

This stage involves data cleaning, feature normalization, and dividing the dataset into training and testing data. These pre-processing techniques are important to ensure that the data used in model training is of high quality and representative. Data cleaning removes noise and inconsistencies, feature normalization ensures each feature has the same scale so that the model is unbiased, and dataset division allows for more accurate evaluation of model performance.

Algorithm Selection

Several machine learning algorithms were tested, including Logistic Regression, SVM with various kernels, CNN, and MLP, based on literature showing good performance in health problem classification. Logistic Regression was chosen for its simplicity and clear interpretation on tabular data. SVM is effective for small to medium datasets and non-linear data. MLP is flexible in capturing complex relationships, while CNN, although usually for image data, is considered for spatial patterns on tabular data.

Evaluation of Results

The models were evaluated with test data to measure the accuracy and computation time of each algorithm. Accuracy is calculated from the percentage of correct predictions, while computation time is in seconds. Confusion Matrix is used to calculate Precision, Recall, and F1 Score, giving an idea of the balance of sensitivity and specificity. The evaluation results are visualized with bar charts and tables comparing accuracy, precision, recall, F1 Score, and computation time, to ensure the selected model is accurate and efficient in detecting stunting.

RESULTS

Logistik Regression

Based on Table 1, the Logistic Regression model achieved an accuracy of 76%. Although the accuracy is relatively lower than some of the other algorithms tested, Logistic Regression has advantages in terms of interpretability and relatively fast computational speed, with a computational time of only 1.789 seconds. This

* Corresponding author

Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13967
e-ISSN : 2541-2019
p-ISSN : 2541-044X

model is less effective in situations where the data has a strong non-linear relationship, which can reduce the accuracy of the predictions as shown in Fig. 2. Fig. 2 shows that Logistic Regression fails to capture complex non-linear patterns in the data, resulting in lower accuracy than CNN, confirming its limitations in dealing with non-linear relationships.

Table 1Logistic Regression Performance

Logistic Regression	**Accuracy**	**Recall**	**F1 Score**	**Precision**
	76%	76%	74%	74%
Time	1.789s			

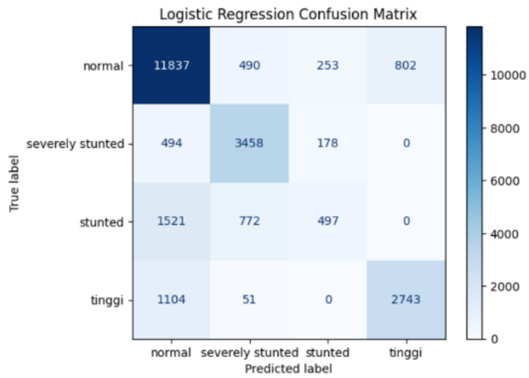

Fig. 2 Confusion Matrix Logistic Regression

Support Vector Machine (SVM)

In this study, SVM with RBF (Radial Basis Function) kernel showed the best performance with 98% accuracy and 54.3 seconds computation time as in Table 2. SVM with polynomial kernel also gives good results with 88% accuracy, but requires longer computation time compared to RBF kernel.

Table 2. Experiment Support Vector Machine (SVM) Performance

	Linear	**Poly**	**RBF**	**Sigmoid**
Accuracy	78%	88%	98%	51%
Recall	78%	88%	98%	51%
F1 Score	77%	87%	98%	48%
Precision	77%	88%	98%	48%
Waktu	214.9s	208.6s	54.3s	459.8s

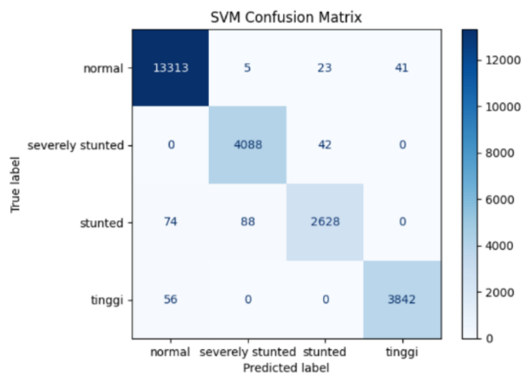

Fig. 3 Confusion Matrix SVM

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) with softmax activation function configuration, Adam optimizer, and sparse_categorical_crossentropy loss function. The model was trained for 10 epochs using relevant datasets. Experimental results show that the model achieved 98% accuracy with a computation time of 81.4 seconds in Table 3. CNN is very effective in handling data with complex and visual features, but its training and computation time are higher than other algorithms in this study. Convolutional Neural Network (CNN) excels in handling data with complex and visual features. In this study, CNN achieved an accuracy of 98%, demonstrating its effectiveness in complex pattern recognition as in Fig. 4 and Fig. 5. CNNs consist of convolution and pooling layers that enable detailed feature extraction, which is very useful in image and video analysis. However, the main drawback of

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13967
e-ISSN : 2541-2019
p-ISSN : 2541-044X

CNNs is their high training and computation time. The intensive training process that requires large computational resources can be a bottleneck, especially in hardware-constrained environments.

Table 3. Experiment Convolutional Neural Network (CNN)

CNN	Accuracy	Recall	F1 Score	Precision
	98%	98%	98%	98%
Time	81.4s			

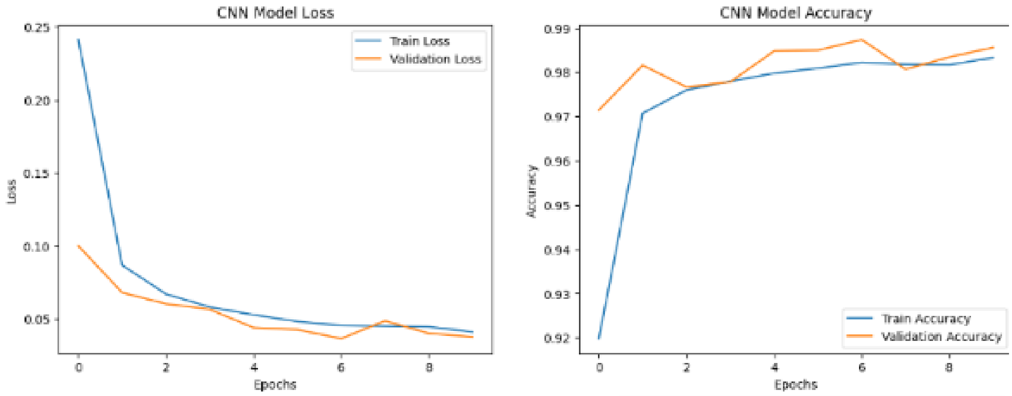

Fig. 4 Loss and Accuracy Curves for Convolutional Neural Network (CNN) Model

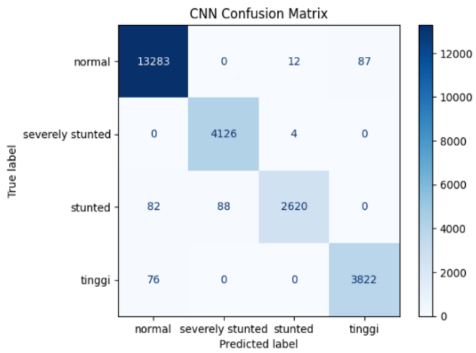

Fig. 5 Confusion Matrix CNN

Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) with the same configuration, namely softmax activation function, Adam optimizer, and sparse_categorical_crossentropy loss function, and trained for 10 epochs. As a result, the MLP model achieved 97% accuracy with 134 seconds of computation time as in Table 4. Although the accuracy achieved by MLP is slightly lower than CNN, the computation time required is much higher. This shows that while MLP can be a good alternative for classification tasks with simpler data structures, CNN remains a more efficient and accurate choice for data with more complex and visual features. Multilayer Perceptron (MLP) also performed well with 97% accuracy in this study Fig. 6 and Fig. 7. The advantages of MLP include its ability to handle data with simpler structures and flexibility in classification and regression tasks. MLP uses the backpropagation method to adjust weights based on prediction error, which allows the model to learn from complex data. However, the disadvantage of MLP is the higher computation time compared to SVM and Logistic Regression. Although effective, MLP may be less efficient for applications that require a quick response or have computational resource constraints.

Table 4. Experiment Multilayer Perceptron (MLP)

MLP	Accuracy	Recall	F1 Score	Precision
	97%	97%	97%	97%
Time	134s			

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13967
e-ISSN : 2541-2019
p-ISSN : 2541-044X

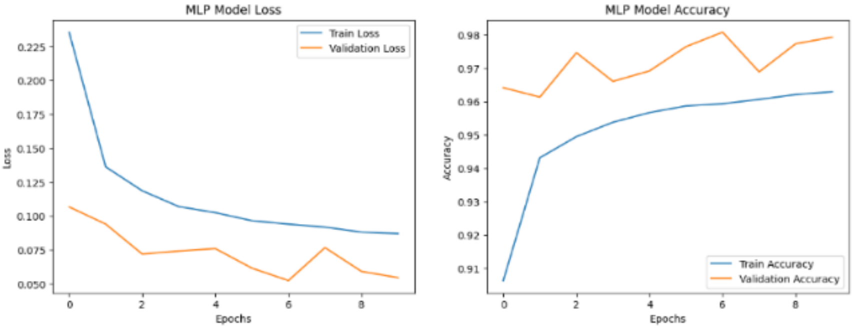

Fig. 6 Loss and Accuracy Curve for Multilayer Perceptron (MLP) Model

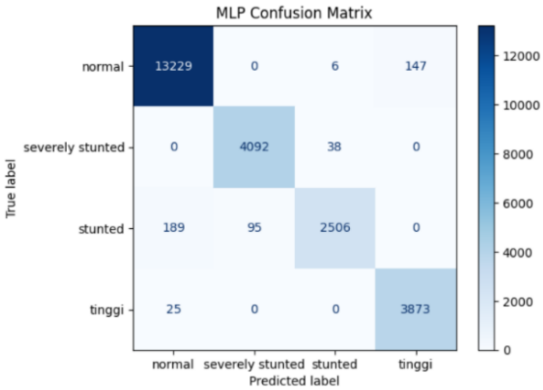

Fig. 7 Confusion Matrix MLP

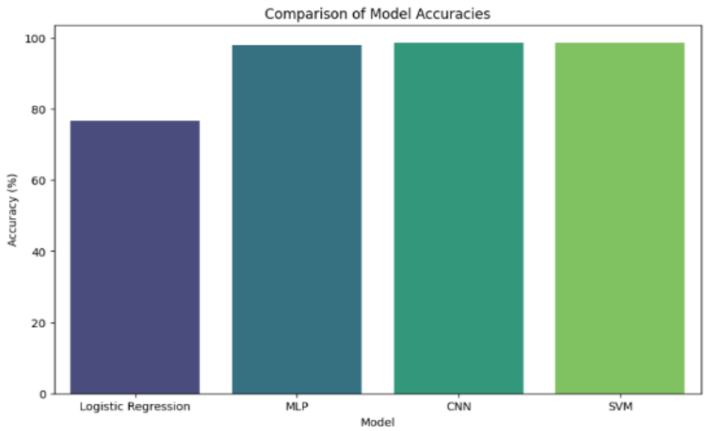

Fig. 8 Machine Learning Model Accuracy Comparison

Table 5
Comparison Results based on chooosed algorithm for stunting detection

	Logistik Regression	**SVM**	**CNN**	**MLP**
Accuracy	74%	98%	98%	97%
Time	1.8s	53.3s	81.3s	134s

DISCUSSION

Research on the application of machine learning algorithms in the Centing (Cegah Stunting) application shows very promising results for early detection and prevention of stunting in children in Indonesia. The four algorithms tested were Logistic Regression, Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, Convolutional Neural Network (CNN), and Multilayer Perceptron (MLP). The results showed that SVM with RBF kernel and CNN achieved the highest accuracy of 98%. SVM with RBF kernel was chosen as the best model because it has a balance between high accuracy and computation time efficiency (54.348 seconds), making it more practical for mobile applications that have limited resources such as battery and processor. CNN, although equivalent in accuracy, requires longer computation time (81.352 seconds) and more resources, which can increase

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13967
e-ISSN : 2541-2019
p-ISSN : 2541-044X

cost and decrease app responsiveness. Logistic Regression, while superior in terms of interpretation and computational speed (1.789 seconds), has lower accuracy (76%), suggesting that it is less optimal for complex and non-linear data. Meanwhile, MLP performed very well with 97% accuracy, but required longer computation time (134.010 seconds) and careful parameter tuning, which could be an obstacle in practical applications. This study suggests that RBF SVM is the optimal choice for Centing applications as it offers the best balance between accuracy, efficiency, and scalability. For future research, the use of automatic parameter optimization methods such as Grid Search or Random Search, and the exploration of other algorithms such as Decision Tree or Random Forest are suggested. In conclusion, the Centing application with SVM RBF model has great potential to support early detection of stunting and improve child health in Indonesia.

CONCLUSION

From the experimental results, SVM with RBF kernel and CNN achieved the highest accuracy (98%) in detecting stunting, with RBF SVM excelling in computational time efficiency. MLP also provides high accuracy but with longer computation time, while Logistic Regression, although fast and easy to interpret, has lower accuracy. SVM RBF was chosen as the best model for the Centing application due to the combination of high accuracy and computational efficiency, essential for applications that must provide fast and accurate detection in the field with limited resources. The addition of attributes such as poverty status can improve the accuracy of the model, but requires effective data management. Each model has advantages and disadvantages in terms of cost and resources: SVM RBF is a good choice for limited resources, CNN is suitable for scenarios that require absolute accuracy, while MLP and Logistic Regression are more suitable for specific needs. The Centing app can be optimized based on field conditions to support effective early detection of stunting across Indonesia. Based on these conclusions, there are several recommendations for future research to expand understanding and improve the quality of stunting detection. It is recommended to add attributes of poverty status, access to local health services, and maternal health status during pregnancy in the analysis to improve the accuracy of the model. The implementation of these suggestions is expected to improve the accuracy, efficiency, and practical application of the stunting detection system in future research.

REFERENCES

Aurima, J., Susaldi, S., Agustina, N., Masturoh, A., Rahmawati, R., & Tresiana Monika Madhe, M. (2021). Faktor-Faktor yang Berhubungan dengan Kejadian Stunting pada Balita di Indonesia. *Open Access Jakarta Journal of Health Sciences*, *1*(2), 43–48. https://doi.org/10.53801/oajjhs.v1i3.23

Banurea, M., Betaria Hutagaol, D., & Sihombing, O. (2023). Klasifikasi Penyakit Stunting Dengan Menggunakan Algoritma Support Vector Machine Dan Random Forest. *Jurnal TEKINKOM*, *6*(2), 540–549. https://doi.org/10.37600/tekinkom.v6i2.927

Fahri, A., & Ramdhani, Y. (2023). Visualisasi Data dan Penerapan Machine Learning Menggunakan Decision Tree Untuk Keputusan Layanan Kesehatan COVID-19. *Jurnal Tekno Kompak*, *17*(2), 50. https://doi.org/10.33365/jtk.v17i2.2438

Gunawan, Muhammad Ichsan Sugiarto, D., & Mardianto, I. (2020). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Seacrh pada Algoritma Logistic Regression. *Jurnal Edukasi Dan Penelitian Informatika*, *6*(3), 280–284.

Hakim, A. R., Atmaja, D. M. U., Tugiman, T., & Basri, A. (2023). Sosialisasi Aplikasi Pelayanan Medis Penyakit Herpes Menggunakan Teknologi Machine Learning. *JMM (Jurnal Masyarakat Mandiri)*, *7*(2), 1633. https://doi.org/10.31764/jmm.v7i2.13935

Hamzah, S. R., & B, H. (2020). Gerakan Pencegahan Stunting Melalui Edukasi pada Masyarakat di Desa Muntoi Kabupaten Bolaang Mongondow. *JPKMI (Jurnal Pengabdian Kepada Masyarakat Indonesia)*, *1*(4), 229–235. https://doi.org/10.36596/jpkmi.v1i4.95

Nugroho, M. R., Sasongko, R. N., & Kristiawan, M. (2021). Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, *5*(2), 2269–2276. https://doi.org/10.31004/obsesi.v5i2.1169

Pitayanti, A., Mulyati, S. B., & Umam, F. N. (2022). Deteksi Dini Cegah Stunting (“ Deni Cheting”) Pada Balita Di Posyandu Krajan Ii. *Jurnal Bhakti Civitas Akademika*, *5*(1), 7. https://doi.org/10.56586/jbca.v5i1.167

Rahman, H., Rahmah, M., & Saribulan, N. (2023). Upaya Penanganan Stunting Di Indonesia. *Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK)*, *VIII*(01), 44–59.

Sanhaji, G., Febrianti, A., & Teknik, F. (2024). Aplikasi DIATECT Untuk Prediksi Penyakit Diabetes

* Corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Menggunakan SVM Berbasis Web. *Jurnal TEKNO KOMPAK*, *18*(1), 150–163.

Sayyidin, M., #1, H., & Muhimmah, I. (2024). Aplikasi Pendeteksi Tingkat Kematangan Pepaya menggunakan Metode Convolutional Neural Network (CNN) Berbasis Android. *Jurnal Edukasi Dan Penelitian Informatika*, *10*(1), 162–170.

Sugianto, M. A. (2021). Analisis Kebijakan Pencegahan Dan Penanggulangan Stunting Di Indonesia: Dengan Pendekatan What Is The Problem Represented To Be? *Jurnal EMBISS*, *1*(3), 197–209. https://www.embiss.com/index.php/embiss/article/view/28

Wahyuni, D., & Fithriyana, R. (2020). Pengaruh Sosial Ekonomi Dengan Kejadian Stunting Pada Balita Di Desa Kualu Tambang Kampar. *PREPOTIF : Jurnal Kesehatan Masyarakat*, *4*(1), 20–26. https://doi.org/10.31004/prepotif.v4i1.539

Widodo, I. E., Handojo, A., & Halim, S. (2020). Aplikasi Pemetaan Penyakit Demam Berdarah di Surabaya dengan Metode Neural Network Multilayer Perceptron. *Jurnal Infra*. http://publication.petra.ac.id/index.php/teknik-informatika/article/view/9764

Wulandari, H. W., & Kusumastuti, I. (2020). Pengaruh Peran Bidan, Peran Kader, Dukungan Keluarga dan Motivasi Ibu terhadap Perilaku Ibu dalam Pencegahan Stunting pada Balitanya. *Jurnal Ilmiah Kesehatan*, *19*(02), 73–80. https://doi.org/10.33221/jikes.v19i02.548

Zulfikar Lating, Mariene Wiwin Dolang, Epi Dusra, Hamka Hamka, & Wa Ode Satriawati Saendrayani. (2023). Analisis Manajemen Kejadian Stunting pada Balita di Desa Waesamu Tahun 2023. *Jurnal Medika Husada*, *3*(2), 21–30. https://doi.org/10.59744/jumeha.v3i2.44

Zurhayati, Z., & Hidayah, N. (2022). Faktor Yang Berhubungan Dengan Kejadian Stunting Pada Balita. *JOMIS (Journal of Midwifery Science)*, *6*(1), 1–10. https://doi.org/10.36341/jomis.v6i1.1730

* Corresponding author