This method, as its name implies, involves a series of cyclic activities. The "sample" phase involves data extraction to obtain a sufficient dataset, allowing significant information to be obtained without becoming overly complex for further processing[5]. The "explore" phase is conducted to delve into the data, searching for trends and anomalies to gain a deeper understanding. "Modify" encompasses the modification of data by creating, selecting, and transforming variables for the modeling process. The "model" phase entails creating a model from the data by automatically seeking combinations that can be used for predictions[10]. "Asses" involves evaluating the patterns found to determine their usefulness and reliability. In general, the block diagram of the proposed system can be seen in Figure 2.

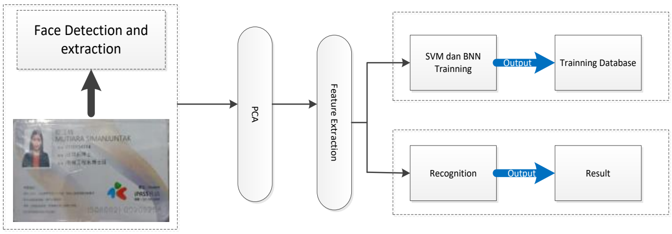

Figure 2. Blok Diagram Sistem Pengenalan Wajah PCA + SVM & BNN

2.1. Dataset

The dataset consists of 324 images in the primary dataset format obtained directly from the research location, namely the Student Affairs Department of AMIK Universal. This dataset comprises files of student identity cards with JPEG extensions (Figure 3).

Figure 3. Image Pra Processing

Subsequently, the images of these Student Identity Cards (KTM) are cropped to capture only the facial area. After cropping, the process transforms the RGB color images into black and white, and further processing adjusts their dimensions and resolution. Table 1 provides detailed explanations of the features of images with specific sizes.

Table 1. Features of Images in Detail

Fitur	Ukuran
Dimension	28 x 28
Weight	28 pixels
Height	28 pixels
Horizontal Resolution	64 dpi
Vertical Resolution	64 dpi
Bit Depth	24

2.2. Principle Component Analysis (PCA

Principle Component Analysis (PCA) is a data analysis technique employed in statistics and is currently widely used in data science. Its objective is to condense large-scale multivariate data tables into smaller sets of variables, commonly referred to as dimensional reduction. The stages of PCA can be seen in figure 4:

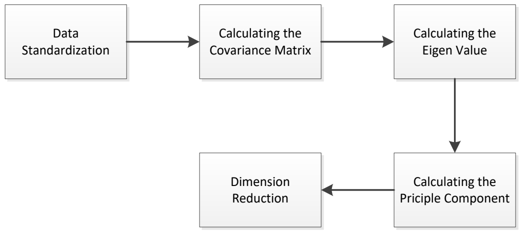

Figure 4. PCA Process Stages

2.3. Algoritma Backpropagation Neural Network (BNN)

In computer vision, robust classification methods are essential to achieve high levels of recognition within constrained time and computational resources. One such method is the Backpropagation Neural Network (BNN). BNN classification is widely utilized for training neural networks due to its simplicity, efficiency in calculating gradient descent, and ease of implementation. Determining neuron size, sample quantity, and weights poses a challenge in the BNN process, and fine-tuning the output neuron is crucial.

BNNs consist of an input layer, one or more hidden layers, and an output layer. The number of neurons in each layer and the connections between them determine the network's architecture. During the forward pass, input data is fed through the network, and activations are computed layer by layer until the output layer is reached. The output is compared to the true labels using a loss function, such as mean squared error for regression tasks or cross-entropy for classification tasks.

In the backward pass (backpropagation), the gradient of the loss function with respect to each weight is calculated by applying the chain rule. Gradients are propagated backward from the output layer to the input layer, adjusting weights to minimize the loss. This process is repeated iteratively, and the algorithm converges to a set of weights that minimize the loss function.

Non-linear activation functions, such as Sigmoid, Tanh, and ReLU, are applied at each neuron to introduce non-linearity, allowing the network to learn complex patterns. The choice of activation function can impact the network's performance and convergence speed. Hyperparameters such as learning rate, number of hidden layers, number of neurons per layer, batch size, and epochs need to be carefully selected and tuned. Grid search, random search, and more sophisticated methods like Bayesian optimization can be used for hyperparameter tuning.

Overfitting is a common challenge in neural networks, where the model performs well on training data but poorly on unseen data. Regularization techniques such as dropout, L1/L2 regularization, and data augmentation can help prevent overfitting. While gradient descent is a fundamental optimization technique, variants like Stochastic Gradient Descent (SGD), Adam, RMSprop, and AdaGrad are often used to improve convergence speed and stability.

Determining the optimal network architecture, such as the number of layers and neurons, requires experimentation and domain knowledge. Proper initialization of weights, usually with techniques like Xavier or He initialization, is crucial for effective training. Monitoring training and validation loss can help detect overfitting and underfitting, guiding the tuning process.

BNNs are widely used in various computer vision tasks due to their effectiveness in learning from visual data. In image classification, BNNs can classify images into predefined categories, such as recognizing objects in photographs or medical images. Beyond classification, BNNs can also be used in object detection frameworks like YOLO (You Only Look Once) and R-CNN (Region-based Convolutional Neural Networks) to locate objects within images. BNNs can segment images into regions of interest, which is essential for applications like autonomous driving, where identifying road boundaries, vehicles, and pedestrians is critical.

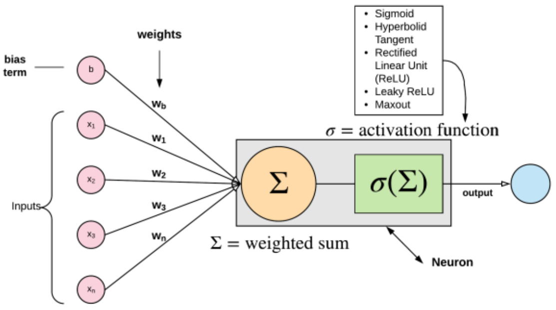

Figure 5. Architecture BNN

2.4. Support Vectore Machine (SVM)

Support Vector Machine (SVM) is a supervised learning method commonly used for classification. SVM operates by searching for the optimal hyperplane, maximizing the distance between classes. Figure 6 illustrates the hyperplane, which functions as a separator between classes.

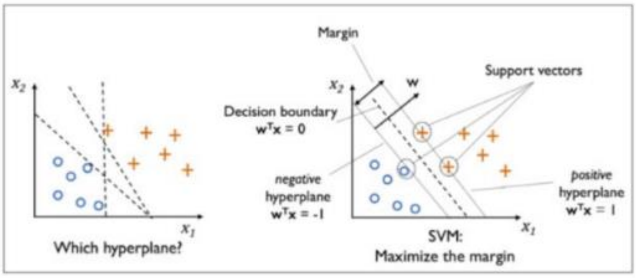

Figure 6. Hyperplane separating the two classes positive (+1) and negative (-1)

The SVM-discovered hyperplane, as depicted in the above image, is positioned in the middle between two classes. This indicates that the distance between the hyperplane and data objects differs for closely positioned (outer) classes marked with empty circles and positives. In SVM, the outermost data objects closest to the hyperplane are referred to as support vectors.

which is the distance between the hyperplane and the nearest data points from each class. These nearest data points, referred to as support vectors, are critical as they define the position and orientation of the hyperplane.

The distance between the hyperplane and these support vectors is crucial because it reflects the robustness of the classification. A larger margin implies a better generalization of the model. In the image mentioned, the classes are represented with empty circles and positive markers, indicating that the outermost data objects (the support vectors) are closest to the hyperplane. These support vectors are the key elements in SVM as they directly influence the hyperplane's location and thus the classification outcome.

2.5. Tools used

To prove the results of this study, the **Orange Data Mining Version 3.2 tool** is used which is equipped with add-ons on Image Analytic which functions as a library to process image data.

III. Result

In this section, the results of experiments that have been carried out will be implemented from the stages of data preparation, modeling and modeling evaluation.

3.1. Data Preparation

Through the import images facility, orange data mining calls data that has been prepared based on gender categories, namely Male and Female, as in the step display in figure 7.

Figure 7. Import Images to call data

Furthermore, the data that has been called can be displayed through the image viewer, so that the results are as shown in figure 8.

Figure 8. Face View using Image Viewe

In order for data to be read through numerical data, orange data mining has an encoding facility involving Image Embedding operators. Figure 9 shows the Embeding Technique using embeded inception 3 architecture.

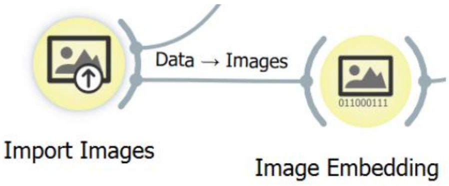

Figure 9. Image Embeding with Inception 3 Architecture

The results of the embeding are processed through the PCA operator to reduce the dimensions of the feature by taking two principle components, so that selected features are obtained as in table 2, figure 10 and figure 11.

Table 2. Selection of Fitur with PCA

	category	image name	image	size	width	height	PC1	PC2
1	Laki Laki	31010140e copy	Laki Laki\31010...	9302	28	28	-14.591	-0.131903
2	Laki Laki	31190001 copy	Laki Laki\31190...	9183	28	28	-19.7074	6.27686
3	Laki Laki	31190002 copy	Laki Laki\31190...	9337	28	28	-17.2239	1.7753
4	Laki Laki	31190004 copy	Laki Laki\31190...	9138	28	28	-18.9142	-0.318818
5	Laki Laki	31190005 copy	Laki Laki\31190...	9269	28	28	-8.85618	1.61283

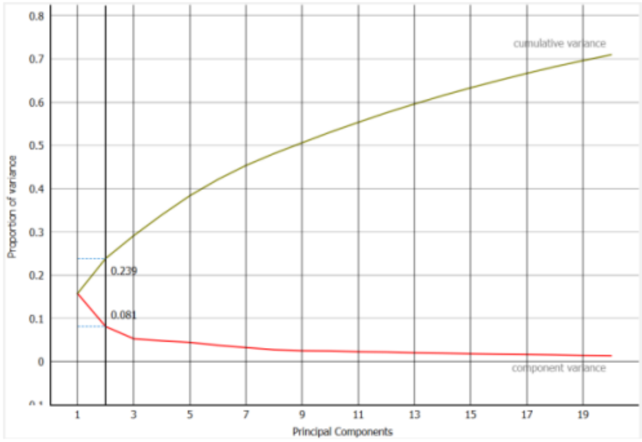

Figure 10. PCA Graphics

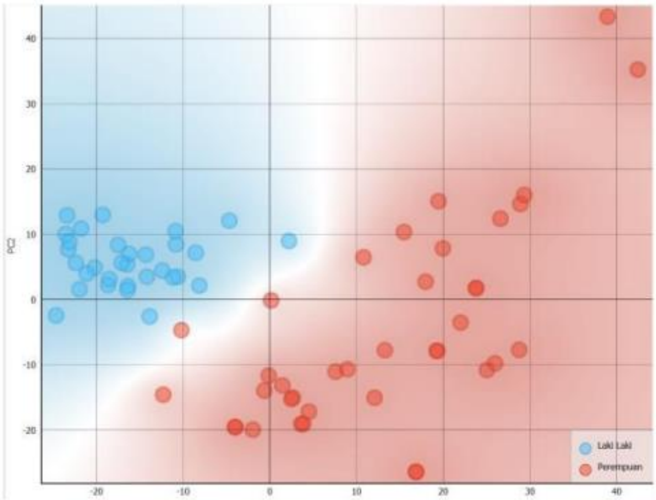

Figure 11. Feature Selection Spread Graph with PCA

3.2. building models

After doing data preparation and the data is ready to be modeled, modeling is carried out with classification algorithms, namely Bacpropagation Neural Network (BNN) and Support Vector Machine (SVM) (figure 12).

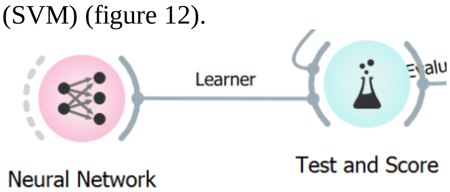

Figure 12. BNN Modeling Design

In the modeling, the parameters used are the number of neurons in layer 10, the activation function using ReLu, and the optimum parameter Adam.

Modeling using the Support Vector Machine classification algorithm can be seen in the design figure 13

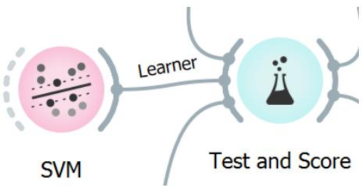

Figure 13. SVM Modeling Design

The parameters covered in SVM itself are SVM Const (C) 1.00, Regression loss epsilon 0.10 with the optimization parameter is numerical tolerance of 0.0010. So that the test results and scores from both modeling can be displayed as in table 3.

Table 3. SVM and BNN Model Development

Model	AUC	CA	F1	Preccission	Recall
SVM	0.848	0.775	0.775	0.775	0.775
ANN	0.861	0.781	0.781	0.783	0.781

From table 3, it can be explained that the Neural Network algorithm model has a better accuracy rate than the SVM algorithm model, which is 78.10%, as well as the Precission and Recall values and F1-score.

3.3. Model Evaluation

Model evaluation is carried out by measuring the performance of the model that has been built through the Confution Matrix table. Confution matrices present tables to measure the performance of classification models. The confution matrix will display and compare the actual value with the value of the model prediction results that can be used to calculate evaluation metrics including Accuracy, Precission, Recall and F1-Score.

Table 4. Confusion Matrix Model SVM

		Predicted		
		Laki-Laki	Perempuan	∑
Actual	Laki-Laki	74.8%	20.1%	150
	Perempuan	25.2%	79.9%	174
	∑	155	169	324

Table 5. Confusion Matrix Model BNN

		Predicted		
		Laki-Laki	Perempuan	∑
Actual	Laki-Laki	74.8%	18.8%	150
	Perempuan	25.2%	81.2%	174
	∑	159	165	324

Based on tables 4 and 5, the male sex prediction value in the SVM algorithm is higher (by 20.1%) than the male sex prediction value in the BNN algorithm. However, the female sex prediction value on the SVM algorithm was smaller (by 79.9%) than the female sex prediction value on the BNN algorithm.

IV. Conclusions

The experimental results on the dataset of facial images from student identity cards, involving PCA dimension reduction technique and modeling with SVM and Neural Network, conclude that there is an improvement in the performance of the developed model compared to the model without PCA for dimension reduction. This indicates that the model design is already quite effective. However, it's important to note that there is still room for maximum improvement in this model. Therefore, future research can focus on enhancing this classification model by implementing deep learning techniques. Additionally, the role of preprocessing steps is crucial, considering the numerous features generated in image objects. Thus, it is expected that applying optimal preprocessing and feature selection techniques will have a positive impact on the performance of the constructed model.

REFERENCE

[1] S. S. Farfade, M. J. Saberian, and L.-J. Li, "Multi-View Face Detection Using Deep Convolutional Neural Networks," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015, pp. 643–650. doi: 10.1145/2671188.2749408.

[2] D. Mishra, K. Rout, S. Mishra, and S. R. Salkuti, "Various object detection algorithms and their comparison," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, p. 330, Jan. 2022, doi: 10.11591/ijeecs.v29.i1.pp330-338.

[3] E. Winarno, I. Amin, H. Februariyanti, P. Adi, W. Hadikurniawati, and M. Anwar, *Attendance System Based on Face Recognition System Using CNN-PCA Method and Real-time Camera*. 2019. doi: 10.1109/ISRITI48646.2019.9034596.

[4] D. Ade, U. Hayati, T. Hartati, S. L. Manikari, F. Afandi, and S. I. Cirebon, "Jurnal Pengenalan Wajah menggunakan Principle Component Analysis (PCA) dengan Model Algoritma Machine Learning untuk Mengidentifikasi Jenis Kelamin pada Kartu Identitas Mahasiswa Face Recognition using Principle Component Analysis (PCA) with Machine L," *J. Pengenalan Wajah menggunakan Princ. Compon. Anal. (PCA) dengan Model Algoritm. Mach. Learn. untuk Mengidentifikasi Jenis Kelamin pada Kartu Identitas Mhs. Face Recognit. using Princ. Compon. Anal. (PCA) with Mach. L*, vol. 5, pp. 219–224, 2023.

[5] Y. N. FUADAH, I. D. UBAIDULLAH, N. IBRAHIM, F. F. TALININGSING, N. K. SY, and M. A. PRAMUDITHO, "Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 10, no. 3, p. 728, Jul. 2022, doi: 10.26760/elkomika.v10i3.728.

[6] D. A. Kurnia, U. Hayati, T. Hartati, S. L. Manikari, and F. Afandi, "Pengenalan Wajah menggunakan Principle Component Analysis (PCA) dengan Model Algoritma

Machine Learning untuk Mengidentifikasi Jenis Kelamin pada Kartu Identitas Mahasiswa," *TEMATIK*, vol. 9, no. 2, pp. 219–224, Dec. 2022, doi: 10.38204/tematik.v9i2.1029.

[7] R. Rosnelly, M. S. Simanjuntak, A. Clinton Sitepu, M. Azhari, S. Kosasi, and Husen, "Face Recognition Using Eigenface Algorithm on Laptop Camera," in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, 2020, pp. 1–4. doi: 10.1109/CITSM50537.2020.9268907.

[8] M. SIMANJUNTAK, "Identifikasi Tanda Tangan menggunakan Metode Fitur Ekstrasi Biner dan K Nearest Neighbor," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 12, p. 191, Mar. 2021, doi: 10.22303/csrid.12.3.2020.191-200.

[9] S. Singh and S. V. A. V. Prasad, "Techniques and challenges of face recognition: A critical review," in *Procedia Computer Science*, 2018, vol. 143, pp. 536–543. doi: 10.1016/j.procs.2018.10.427.

[10] I. F. Yuliati, S. Wulandary, and P. R. Sihombing, "Penerapan Metode Support Vector Machine (SVM) dan Backpropagation Neural Network (BPNN) dalam Pengklasifikasian Pasangan Usia Subur di Jawa Barat," *J. Stat. dan Apl.*, vol. 4, no. 1, pp. 23–34, 2020.