Journal of Applied Engineering and Technological Science
Vol 5(1) 2023: 542-556

A MACHINE LEARNING MODEL FOR DETERMINATION OF GENDER UTILIZING HYBRID CLASSIFIERS

Dewi Nasien[1*], M. Hasmil Adiya[2], Yusnita Rahayu[3], Dahliyusmanto[4], Erlin[5], Devi Willieam Anggara[6]
Department of Informatic Engineering, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, 28127, Indonesia[125]
Department of Electrical Engineering, Universitas Riau, Pekanbaru, 28293, Indonesia[34]
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia[6]
dewinasien@lecturer.pelitaindonesia.ac.id

Received : 14 March 2023, Revised: 28 November 2023, Accepted : 01 December 2023
Corresponding Author

ABSTRACT
This study addresses the challenge of identifying gender in forensic anthropology when faced with incomplete, burned, or damaged skeletal remains. The objective is to utilize the pelvis and femur, established as reliable gender indicators, for accurate identification. Employing measurements of the subpubic angle of the pelvis and femur angles, a principal component analysis (PCA) method is applied to create two attributes for machine learning model input. The study incorporates a hybrid machine learning system, combining an Artificial Neural Network (ANN) and Support Vector Machine (SVM) design. Testing with acquired data yields an 83.33% accuracy, classified as "good" based on the Area Under the Curve (AUC) from the confusion matrix. The implications of this research extend to both theoretical and practical domains, providing a method for accurate gender identification in challenging forensic scenarios. The contribution lies in offering a reliable and innovative approach that can be applied to enhance gender determination practices in forensic anthropology, thereby advancing both theory and real-world applications.).
Keywords: *ANN-SVM, PCA, Pelvic, Femur, Gender Determination*

1. Introduction

Forensic Anthropology is a sub-discipline of scientific study that involves cases of humanitarian, medico-legal, or forensic interest (Alfsdotter, 2021). Forensic Anthropology is currently one of the fastest-growing sub-disciplines of biological anthropologies, that is focusing on the identification of skeletal remains or skeletons (Afrianty et al., 2015). Skeletal remains are usually exposed to various factors that could influence their condition before it is found, such as burnt, damaged, broken, and bones missing, which led to scientists finding it hard to identify the skeletal remains. The process of identification of skeletal remains by an expert is using DNA (Deoxyribose Nucleic Acid) analysis in the laboratory (Cattaneo, 2007; Hairuddin et al., 2021). However, the analysis of DNA has some disadvantages such as the skeletal remains cannot be extracted if the skeleton is burnt or damaged, thus the sample cannot provide data for the analysis process (Cattaneo, 2007).

There are four important parameters in forensic anthropology for identifying skeletal remains, namely the determination of age, stature, sex, and ancestry (race) (Afrianty et al., 2022; Yaşar Işcan & Olivera, 2000). The determination of sex is the main and most important step in the identification of the human skeleton (Afrianty et al., 2015; Yaşar Işcan & Olivera, 2000). In the field of forensic anthropology, sex determination is part of the pattern recognition process that aims to ascertain whether the skeleton is from a male or female individual. Knowledge of sex is essential in identifying additional information related to race, age and body shape (Nasien et al, 2021).

Currently, the best method to identify the sex in human skeletal remains is based on the presence of a well-preserved pelvis, but in reality, most of it is usually found broken. Another element that can be used to identify the human sex in remains is the femur. The femur has previously been used and proven to be more accurate than the pelvis to identify sex in skeletal remains from a bio-archaeological and forensic standpoint (Curate et al., 2016).

The pelvis is agreed to be the most frequently used part of the human skeletal remains as an attribute to identify the sex of the deceased (Afrianty et al., 2022; Curate et al., 2016; Darmawan et al., 2021; Phenice, 1969). Initially, the skull has been considered the second-best indicator for the identification of sex in skeletal remains, however, the latest research has shown that when the pelvis is missing or broken, the current most frequently used part for assessing the sex in remains is the postcranial elements (Spradley & Jantz, 2011). Postcranial elements used are the tarsals (Navega et al., 2015), long bones [(Christensen, Passalacqua, et al., 2014; Holland, 1991; Kranioti et al., 2009; Navega et al., 2015; Spradley & Jantz, 2011), the sternum (Macaluso & Lucena, 2014a), the scapula and clavicle (Papaioannou et al., 2012), the vertebrae (Gama et al., 2015), and the femur (Curate et al., 2016). Among all parts of the human skeleton, the pelvis is considered to be the part that shows the most differences between men and women. This is due to sexual dimorphism that is strongly associated with the selective pressures of upright walking (bipedalism) and labor. Adaptive differences in pelvic structure are also influenced by sexual selection. In addition, pelvic anatomy is affected by developmental plasticity, which can be caused by ecological, climatic, and nutritional factors, as well as by neutral demographic processes (Coelho & Curate, 2019).

The next element used in this study is the femur. Besides the pelvic bone, the femur can also be used in determining sex because long bones, including the femur, show clear sex differences. Analysis of the overall length and size of the long bones generally facilitates sex determination. In general, the male femur is longer, larger, and has more obvious muscle attachments than the female femur. The size of the epiphysis, particularly on the femur, has been shown to be accurate in determining sex efficiently (Santosh et al., 2022). The femur is the strongest and heaviest bone in the skeleton thus it is often used in archaeological and forensic contexts (Macaluso & Lucena, 2014b; Papaioannou et al., 2012). Few dimensions of the femur such as femoral length, bicondylar breadth, and femoral head have been used for assessing sex in unknown skeletal individuals (Holland, 1991; Kranioti et al., 2009; Phenice, 1969). The single best femoral measurement of sex is probably based on the head diameter of the femoral (Stewart, 1979), but recent studies also demonstrated other proximal femur dimensions capacities, such as the Femoral Neck Width (FNW) or Femoral Neck Axis Length (FNAL) for assessing sex (Mays, 1992; White et al., 2012) and ancestry attribution (Mays, 1992; Stewart, 1979).

The method used in this study is hybrid machine learning. There are two methods of machine learning that combine to be a hybrid model, namely ANN-SVM. ANNs are usually robust in handling complex and non-linear patterns, while SVMs can be effective in handling classification problems with clear decision boundaries. Various types of research have proven the efficiency of hybridizing ANN with SVM to make classification more effective and have a faster process (Meeusen et al., 2015; Stojanowski, 1999). The reason for using a hybrid method instead of the single technique is that the hybrid proved that it gives better results with a more accurate comparison to the single method (Christensen, Leslie, et al., 2014; Lee et al., 2013). The Hybrid ANN SVM model proved to be faster in the testing phase and was able to improve classification performance by reducing the error rate. The combined approach of ANN and SVM classifiers integrated in one hybrid system has been successfully applied to classify the location of damaged parts in aircraft gas turbine engines, perform soil moisture content prediction, and handle visual and recognition tasks in the context of robotic swarm use (Al-Boeridi et al, 2015).

While the pelvis and femur are valuable tools for gender determination, relying solely on these bones can present several limitations that could impact the accuracy and reliability of the analysis.If the training data used to develop machine learning models for gender determination is biased towards a specific population or set of criteria, the model may not perform accurately when applied to individuals from different backgrounds.Trauma or diseases affecting the pelvis or femur can significantly alter their morphology, making gender determination based on these bones challenging or even impossible.In cases of incomplete skeletal remains, where the pelvis or femur is missing or severely damaged, gender determination may be hindered or require additional sources of information.

Based on the above background, this study aims to develop an ANN-SVM hybrid model in identifying gender using pelvis and femur dataset for classification. The hybrid approach is chosen because it can utilize the advantages of both models, namely the ability of Artificial Neural

Networks (ANN) in handling complex data and Support Vector Machine (SVM) in handling classification problems. By combining the two, it is expected that this research can improve the classification accuracy of sex identification from the femur dataset by utilizing the advantages of each model to achieve better results. In addition, this research also aims to explore the potential of pelvic and femur measurements as reliable indicators for sex determination, although it must be recognized that there are limitations. Therefore, this research will detail an attempt to explore the extent to which these two elements can be a reliable source of information in the context of sex identification in human skeletal remains.

2. Literature Review

A. Artificial Neural Network (ANN)

An artificial Neural Network (ANN) is described as an array of neurons that are connected to each other. ANN is inspired by the learning process of the human brain biologically (Anil K & Jianchang, 1996). ANN has been used widely for the practical application of artificial intelligence, for example, classification, function approximation, pattern recognition, etc. (Rabunal & Dorado, 2005). Other than that, ANN has been accurately used for fault detection in machines (Erlin et al., 2022b; Fayyad et al., 1996), and lately, researchers have also used ANN to diagnose failures in the gearbox (Anil K & Jianchang, 1996; Paya et al., 1997; Rabunal & Dorado, 2005).

ANN contains some information on the processing system modeled from the human brain's neural networks, which consists of simple elements that work in parallel. The purpose is to solve complicated problems based on internal and external information available (Afrianty et al., 2015). The ANN's ability to predict outcomes successfully depends on selecting proper weights while it processes the training. During the learning and training processes, inputs and targets are connected (Afrianty et al., 2015).

Artificial Neural Networks (ANN) draw inspiration from the brain's neural networks. They are useful in pattern recognition, flaw detection in machinery, and classification because of their capacity to model complex relationships. ANNs could detect patterns and correlations among skeletal features in the context of determining gender from skeletal remains, which would help classify these features and identify gender-specific traits.

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the famous machine learning that has supervised algorithms. The supervised algorithm in SVM is used for the regression and classification problems by providing numeric data (Belousov et al., 2002).

The SVM is one of the most well-known and used machine learning. The SVM is a major technique for classifying linear and non-linear data. It uses a non-linear mapping to combine the original training data into a higher dimension. A new dimension of the SVM searches for linear optimal separating hyperplane. The suitable non-linear mapping can always be separated by a hyperplane to an adequately great extent. This hyperplane is found by SVM using margins and support vectors themselves. SVM performs classification task b, maximizing the distinct margin for both classes while the error classification is minimized (Shylaja & Muralidharan, 2019).

The most crucial feature of SVM is the solution that is represented by using the examples belonging to a subset of the original data train for both classification and regression, and this process can decrease cost reduction from a computational point of view (Franchini et al., 2023).

Support Vector Machines (SVM), which are well-known for their classification skills, are excellent at differentiating across classes by building the best possible separating hyperplanes. SVM may help distinguish important traits that clearly differentiate male and female traits when determining gender from skeletal remains.

C. Hybrid ANN SVM

The combination of Support Vector Machines (SVM) and Artificial Neural Networks (ANN) in a hybrid technique provides unique benefits that greatly improve classification accuracy. ANNs are excellent at identifying complex patterns and correlations in data, which makes them perfect for identifying subtle traits in the identification of gender from skeletal remains.

SVM, on the other hand, specializes in finding the best separating hyperplanes, which is especially helpful in cases when classes are well-defined. The hybrid technique guarantees less overfitting and enhanced generalization by combining these strengths. In order to reduce noise and improve robustness, the combined model makes use of ANNs' capacity to represent intricate relationships and SVM's emphasis on critical support vectors. Furthermore, this combination takes advantage of ensemble learning, providing different viewpoints on the data and addressing non-linear complexity found in problems involving gender determination.

In the realm of forensic anthropology, machine learning techniques are harnessed to discern gender-specific skeletal indicators. Through training on datasets containing known gender-specific bone traits, both ANN and SVM learn to categorize these features, potentially culminating in a more precise and reliable gender detection methodology for forensic applications. While hurdles exist, the integration of machine learning techniques in forensic anthropology marks a promising stride.

Moreover, this amalgamation benefits from ensemble learning, offering diverse perspectives on data and effectively tackling non-linear complexities inherent in gender determination problems. This combined approach not only promises accuracy in gender identification but also signifies a robust leap towards refining forensic methodologies.

D. Data Management and Processing

Data management and processing are related to multifactorial age estimation. Two of the most frequently encountered issues are redundancy and incomplete or lost fragments of data. When bilateral or paired data is gathered, redundancy is unavoidable. Human anatomy is complex, and the structure of a human body is never entirely symmetrical. However, unless a particular abnormality exists, the left and right sides of a human body should not appear too different from each other.

One common solution to address redundancy and incomplete data involves the use of heuristics and imputation techniques. In cases where bilateral traits are analyzed, a heuristic approach is employed by selecting one side as the main data source. Typically, the left side is chosen, and if it contains incomplete data, the right-side data is used as a substitute in calculations. This heuristic helps streamline the analysis process, making it more manageable. To further address incomplete data, imputation techniques, such as the simple closest neighbor approach, are utilized. In this method, missing values are imputed by replacing them with the values of the closest neighbor ($k = 1$), determined through Jaccard similarity on one-hot encoded data. This approach is a practical way to fill in missing information while preserving the overall structure of the dataset.

The method described simplified the number of skeletal features from 99 to 64, thus minimizing redundancy and dimensionality effectively. According to previous research (Beretta & Santaniello, 2016), this was a favorable way to preserve a dataset's structure. The research described a next-level algorithm that can also be used to reduce imputation error, but the program's downside is it causes significant data distortion. Sexes were pooled in order to increase the available data's age-related variability and volume. Even though there are not many specific considerations when choosing between the methods described earlier, one must remember that in forensic anthropology, sex is identified during casework. Sex-specific models and pooled data models each neutral out the advantages and disadvantages each has to offer, thus canceling out the mis-specifications.

In summary, the challenges of handling incomplete or redundant data in forensic anthropology necessitate the use of heuristics and imputation techniques. The approach of choosing a main data source and imputing missing values based on the closest neighbor's values proves effective in preserving data structure, minimizing dimensionality, and addressing the unique complexities of age estimation in forensic anthropology.

The unqualified data represented 8.18% or 9 data of the total values in the datasets, so the conclusion is only 91.82% or 101 data of the datasets that can be used in the machine learning models. While only 91.82% of the initial dataset is usable due to unqualified data, the remaining portion still offers a substantial foundation for our machine learning models. This curation ensures data quality, mitigating noise and inconsistencies. The refined dataset supports meaningful results by maintaining a balance between capturing essential patterns and excluding potentially biased information. Despite the reduction in size, the 91.82% of usable data enhances the reliability and accuracy of our analyses, emphasizing quality over quantity in machine learning model development.

The 101 data used in the machine learning model will be separated into data train and data test. The Data train is represented as x_train while the data test is represented as x_test. The comparison of the separation is 70:30, which means there will be 71 data trains, and there will be 30 data tests. The 70:30 data split in machine learning, allocating 70% for training and 30% for testing, is a pragmatic choice based on key considerations. It provides sufficient training data for the model to learn patterns and relationships without overfitting, while the reserved 30% enables a robust evaluation on unseen examples. This split strikes a balance in line with industry practices, ensuring both efficient model training and effective evaluation, particularly in resource-limited scenarios. Overall, the 70:30 split is widely accepted as a practical compromise for achieving adequate model training and rigorous evaluation.. Every single data has its target. 0 for male and 1 for female. The target is illustrated in y, and there will be y_train and y_test. In our study, gender objectives are encoded as 0 for male and 1 for female, simplifying the classification process. This binary encoding aligns seamlessly with widely used machine learning methods, enhancing computational efficiency in model training, assessment, and prediction. The uniformity of this encoding improves study findings' interpretability and communicability, making statistical analysis, including metrics like accuracy and precision, more straightforward. The binary representation mirrors the societal conception of gender, adhering to traditional categorization systems and supporting our research goals with a simple and effective framework. This methodology ensures clarity and efficacy in our study, making it accessible to a broad readership.

3. Research Methods

The methodology used in this research can be seen in the methodology flow in figure 1 below.

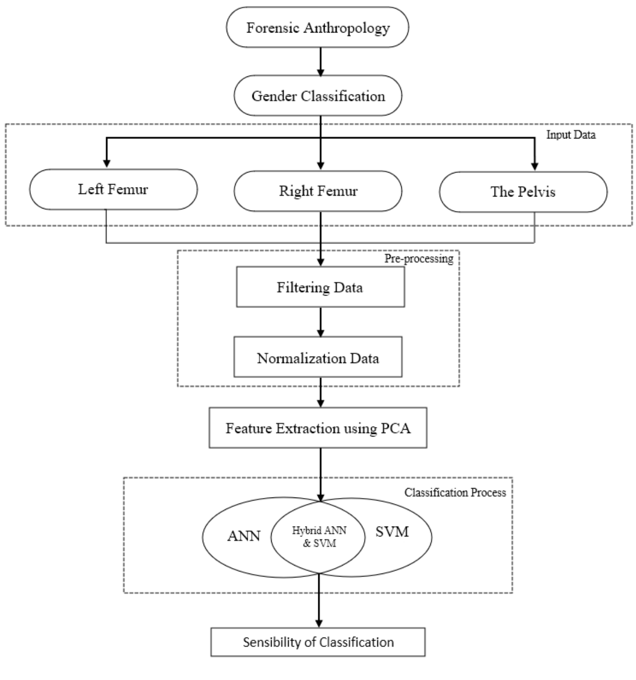

Fig. 1. Methodology Flow

Based on the framework above, the process starts with Forensic Anthropology. In this process, the datasets were separated by their gender and attributes. The gender qualifications are male and female, while the attributes are left and right femur and the pelvis. Then, the gender is classified by the femur and the pelvis. These attributes play a significant role in gender determination due to the distinct biological differences between males and females. For instance, male femurs are typically longer and thicker than female femurs, while female pelvis are broader to accommodate childbirth (Akhlaghi et al., 2019). The measured femur angle is the angle formed between the pubis and the pubic symphysis. As shown in figure 2.

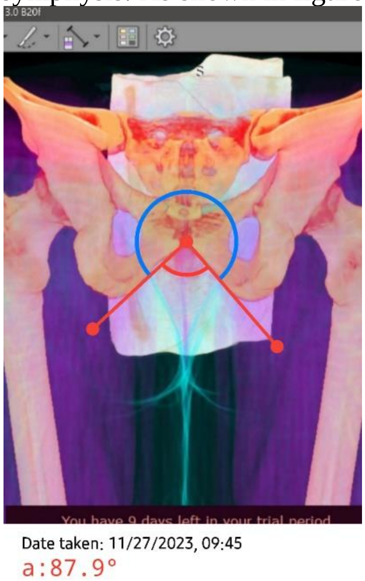

Fig. 2. Pelvic Angle Measurement Results

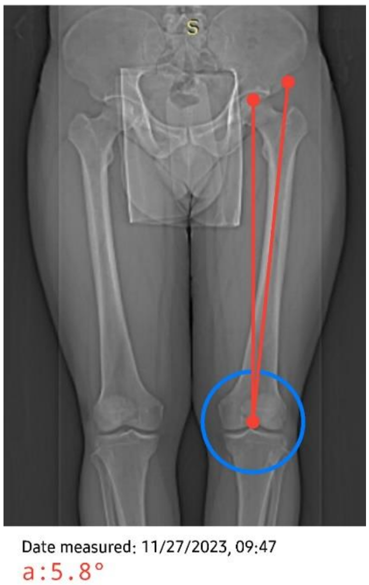

Fig. 3. Femur Angle Measurement Results

In this process, the datasets are ready to be included in the next step. When the datasets are prepared, the attributes, left femur and right femur, and the pelvis are being input data to the machine learning. After the input data, the next step is pre-processing. In pre-processing, there are two stages. The first one is to filter the datasets; the broken or unquantifiable datasets will be deleted from the datasets. Some of dicom file that we have, have blurry images. Excluding blurry images and images that are not fully visualized from your dataset is an important step in ensuring the quality of data used for analysis. This is because these images are likely to contain inaccurate or incomplete information, which could lead to unreliable results. The selected datasets are the only data that will be processed in the machine learning models. After the filtering data process is done, the next stage is normalization data. The filtered datasets will be normalized. The normalization process makes the dataset's values in the range of 0.1 – 0-9. Normalization data makes machine learning models easier to classify the datasets. Normalizing filtered datasets is crucial for optimizing machine learning model efficiency. This process ensures a consistent scale across features, preventing dominance by larger numeric ranges and facilitating faster convergence during training. It enhances the model's generalization by reducing sensitivity to input data variations and mitigates numerical instabilities. Overall, normalization contributes to both interpretability and the effectiveness of machine learning models (Singh & Singh, 2020; Suarez-Alvarez et al., 2012). The next step is Feature Extraction. Feature extraction is a crucial step in machine learning, particularly when dealing with high-dimensional datasets (Gorodetsky & Samoylov, 2008; Shamim & Yogesh, 2021).

A high-dimensional dataset can lead to several challenges, including increased computational complexity, overfitting, and difficulty in data visualization. Principal Component Analysis (PCA) is a popular and effective technique for feature extraction. In this method, the attributes will be combined into two attributes only. The two attributes, x, and y, can be represented in the graphic x and y axes so it can be shown good visualization and make it easier to understand the datasets. This technique increases the number of minority classes by synthesizing data samples while maintaining the number of majority classes (Erlin et al., 2022a). After the feature extraction process, the next step is classification. The classification process employed the Artificial Neural Network (ANN) method and Support Vector Machine (SVM). The model will create a hybrid model, and it includes the ANN model and SVM model. The hybrid model will be created using the estimator library in python. The estimator library allows model A (ANN) combines with model B (SVM). The combination of the two models or hybrid model ANN SVM will be used to classify the datasets of the bones. The hybrid model, created with the \`VotingClassifier\` from the \`sklearn\` library in Python, combines the strengths of the Artificial Neural Network (ANN) and Support Vector Machine (SVM). The ANN captures complex relationships, while the SVM excels in high-dimensional spaces. By leveraging their complementary capabilities, the hybrid model aims to enhance classification accuracy and versatility, providing a robust solution for analyzing bone datasets. The hybrid model, comprising both an Artificial Neural Network (ANN) and a Support Vector Machine (SVM), is applied to classify bone datasets through a collaborative decision-making process. Input attributes, such as features related to bone structure and density, are processed and represented to capture their nuanced relationships. The ANN component excels at learning complex, non-linear mappings, extracting intricate patterns from the high-dimensional data. Concurrently, the SVM establishes robust decision boundaries, particularly effective in handling complex and high-dimensional datasets. The outputs from both models are integrated in a manner defined by the hybrid model architecture, leveraging the unique strengths of each for improved classification accuracy. In predicting gender based on input attributes, the hybrid model considers the combined insights from the ANN and SVM components. The model's ability to discern intricate patterns and establish clear decision boundaries enhances its effectiveness in accurately predicting gender from bone datasets. The collaborative integration of ANN and SVM ensures a comprehensive and nuanced approach to the decision-making process, ultimately contributing to the model's accuracy and reliability in gender classification. The training process was conducted using a combination of software and programming languages that facilitated various stages of the process. Google Colaboratory served as the primary platform for running the training and development of the machine learning model. TensorFlow was instrumental in constructing the model architecture and training it on the prepared datasets. Additionally, scikit-learn provided various tools for data preprocessing, feature engineering, and evaluation of the model's performance. Notably, scikit-learn was used to implement Principal Component Analysis (PCA) for dimensionality reduction and to create a hybrid model that combined multiple machine learning algorithms. Matplotlib, a data visualization library, was employed to create informative plots and graphs that aided in understanding the data and interpreting the model's results.

The dataset used in this study comes from Universiti Teknologi Malaysia. Acquiring the datasets from Universiti Teknologi Malaysia involved a thorough search and selection process to ensure the quality and diversity of the data. The datasets were carefully chosen to represent a wide range of ages and population origins, enhancing the model's robustness and generalizability. The dataset contains images of the pelvis and femur bones. There are 49 male subjects and 52 female subjects, which adds up to a total of 101 images of bones. The pelvis and femur angles will be calculated with Angulus application. There are two angles from the femur; the left femur and the right (Figure 3) and one angle from the pelvis (Figure 4).

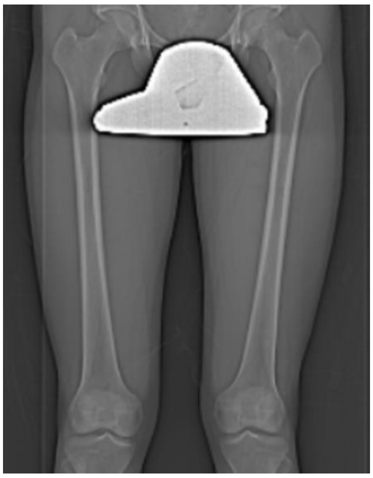

Fig. 3. The Femur

Fig 4. And The Pelvis

Certainly, the information obtained from the displayed bone images is multi-faceted, providing valuable insights into anatomical characteristics crucial for gender determination. The three key pieces of information extracted from the images include the angles of the left femur, right femur, and pelvis. These angles serve as quantitative measures, contributing to the precision and accuracy of the gender classification process.

4. Results and Discussions

The results of the experiment are divided into a few steps, such as Data Collection, Feature Extraction (Shown in Figure 5), Classification, and Implementation. In the Data Collection process, all the data will be normalized and filtered. Normalized and filtered data will be processed in the next step, Feature Extraction.

In Feature Extraction, the dataset will be converted to a simple dataset to make the classification process easier for machine learning and increase the result of the machine learning. The original dataset is in 3 attributes, such as the angle of the left and right femur and the angle from the pelvis. These attributes are converted to be 2 attributes only, with the PCA method in Feature Extraction, which are X and Y axes. In Figure 4, the male and female samples are separated. Red dot for male, and green dot for female. Also, in this Feature Extraction process, the data will be shown their own region. Based on the result of Feature Extraction, the region of the male samples is in value 0 to 3 for the x-axes of the graphic, while the region of the female samples is in value -2 to 0 for the x-axes of the graphic. The next step is classification. In the classification process, the models of hybrid ANN-SVM will be created. Before implementing the models, there will be testing to find the best parameters. The parameters are batch size and epoch; from the testing, the best parameters are 5 for batch size and 10 for the epoch. The values of these parameters will be implemented in the models.

Parameter tuning, encompassing adjustments to batch size and epoch, holds a crucial role in shaping the performance of the hybrid model (KORKMAZ TAN & BORA, 2017; Mohamad et al., 2020). A smaller batch size of 5 facilitates frequent parameter updates during training, enabling the model to capture nuanced patterns in bone datasets. The chosen epoch value of 10 ensures a balanced training approach, allowing the model to learn from the data sufficiently without succumbing to overfitting. The selection of these values likely emerged from an iterative testing process, where the model's performance metrics, including accuracy, precision, recall, and F1-score, were assessed across various parameter combinations. The influence of these tuned parameters on the hybrid model is substantial, contributing to its overall effectiveness. The smaller batch size aids in adapting quickly to data patterns, while the moderate number of epochs strikes a balance between underfitting and overfitting. This careful parameter tuning process enhances the model's accuracy and generalization ability, crucial for its application in real-world forensic anthropology cases (Thomas et al., 2017).

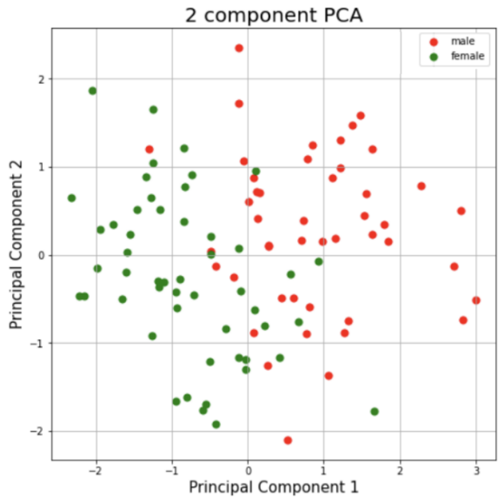

Fig. 5. PCA Feature Extraction Result

The result of the machine learning is described as 0 and 1. The result needs to be compared to the actual data. There are 30 data that have been tested. From the testing data of the machine learning, there are 25 correct data and the others are false. The comparison between the prediction and the actual dataset is plotted in Figure 6.

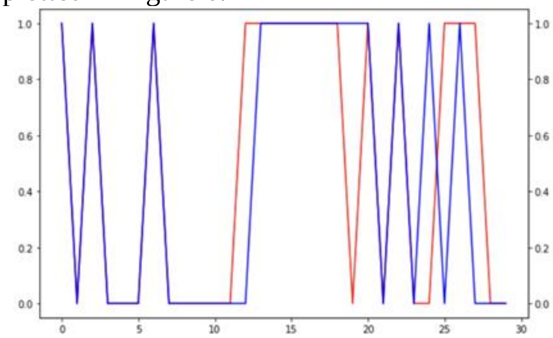

Fig. 6. Comparison Between Actual And Prediction Data

In Figure 5, there are two lines, blue and red. Blue represents the actual data, while a red represents the prediction data. There are some data that are not matched to each other, which means the prediction is false. The last step is calculating the accuracy of the result with a confusion matrix. The confusion matrix is created by combining the actual values (including positive and negative) and the prediction values (including positive and negative). The confusion matrix is depicted in Figure 7.

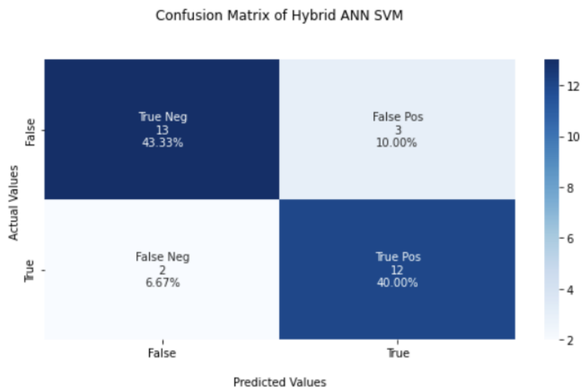

Fig. 7. Confusion Matrix Of The Prediction

From Figure 7, the accuracy is calculated by summing all of the true positive (values 13, percentage 43,33%) and true negative (values 12, percentage 40,00%). The total accuracy is 83,33%. Based on the Area Under Curve (AUC) (Gorunescu, 2011), the model of machine learning is a good classification. The precision is calculated by the sum of all true positive values from the confusion matrix. The values will be divided by the total of true and false positive values. The precision is (12 / (12+3)) * 100% equals to 80%. The recall value is the same with precision, but the division utilizes the sum of true positive and false negative. The recall is (12 / (12+2)) * 100% equals to 85.71%. The last value, specificity, is the value of true negative values divided by true negative plus false positive, which means (13 / (13+3)) * 100% equals 81.25%.

Discussion

While an accuracy value of 83.33% is indicative of a relatively high level of correctness in the model's predictions, it's essential to assess its statistical significance and practical applicability in forensic anthropology cases. The determination of whether this accuracy is considered robust depends on factors such as the complexity of the dataset, the distribution of classes, and the consequences of misclassifications (Benz et al., 2021; Dogan & Tanrikulu, 2013; Fan & Koutris, 2022; Lian et al., 2021).

Achieving an accuracy of 83.33% in gender determination using skeletal remains is a noteworthy accomplishment with profound implications for forensic anthropology. This level of accuracy holds significance in comparison to existing methods, showcasing competitive performance, particularly in the face of challenges posed by incomplete or damaged skeletal remains. The robustness of the proposed approach, leveraging the pelvis and femur within a hybrid machine learning system, enhances the reliability of gender determination, crucial in forensic investigations. The practical impact extends to real-world scenarios, where accurate gender identification is essential for establishing biological profiles of unidentified individuals. The method's success rate serves as a valuable tool for law enforcement agencies and forensic experts, streamlining and improving the efficiency of forensic investigations. The hybrid machine learning system, combining an Artificial Neural Network (ANN) and Support Vector Machine (SVM), offers a complementary and comprehensive approach, contributing to the advancement of gender determination methods in forensic anthropology. Moreover, the achieved accuracy provides a solid foundation for further refinement and potential improvements, promising continuous advancements in the field.

The precision, recall, and specificity values calculated for the model's performance in classifying male and female samples within the bone datasets provide crucial insights. With a precision of 80%, the model demonstrates accuracy in positive predictions, indicating that 80% of instances predicted as male or female are indeed accurate. The recall, or sensitivity, is notably high at 85.71%, emphasizing the model's effectiveness in capturing true positive instances and minimizing the risk of false negatives. Additionally, the specificity value of 81.25% underscores the model's accuracy in negative predictions, essential for preventing false positives in forensic anthropology cases. Collectively, these metrics offer a comprehensive understanding of the model's gender classification capabilities. The high precision, recall, and specificity values signify a balanced and reliable performance, crucial for accurate gender determination in real-world forensic applications where precision is paramount.

Analyzing instances where predictions deviated from actual data offers crucial insights into the model's limitations. False predictions may stem from individual anatomical variations departing from typical patterns, influenced by factors such as bone morphology, pathology, or population-specific differences. Additionally, the model's challenge in capturing subtle variations in bone structures, exacerbated by taphonomic factors like degradation or damage to skeletal remains, contributes to false predictions. These misclassifications emphasize areas for improvement, indicating a necessity for more diverse training datasets, enhanced features, and considerations for population-specific characteristics. Adjustments to model architecture, hyperparameters, and the incorporation of ensemble learning methods could mitigate the impact of overfitting or underfitting. This analysis underscores the intricate nature of gender determination from skeletal remains, guiding future enhancements for a more accurate and robust forensic anthropology tool.

Hybrid ANN-SVM approach emerges as a valuable complementary tool for gender determination, this method's success in achieving an 83.33% accuracy, suggests its potential to enhance accuracy in gender determinations, offering a streamlined identification process for forensic experts. This becomes especially relevant in resource-constrained situations where DNA analysis is impractical or inconclusive. Machine learning model allows for practical implementation in the field, enabling anthropologists to integrate it into their toolkit alongside existing techniques. Ethically, the approach addresses concerns related to privacy and respectful treatment of remains, providing a non-invasive yet accurate means of gender determination. Overall, our study's findings present a promising and versatile method that can help forensic anthropology practices, offering practical solutions in real-world scenarios where DNA analysis is not a viable option.

The reliability and generalizability of our findings are paramount considerations in evaluating the applicability of our hybrid ANN-SVM model to diverse datasets, populations, and forensic scenarios. The achieved 83.33% accuracy underscores the reliability of our model in gender determination from skeletal remains. However, its generalizability to other datasets relies on the representativeness and diversity of the training data. To enhance this, future studies should explore incorporating datasets from varied geographical regions, ethnicities, and time periods. Population-specific considerations and adjustments to the model, such as retraining on specific subsets, can further bolster its effectiveness across different populations. Limitations and biases, including the reliance on specific skeletal features and potential influences of age, health conditions, or cultural practices, must be acknowledged. Additionally, the impact of taphonomic factors on skeletal remains could vary across scenarios, necessitating careful consideration for generalizing the model's effectiveness.

Future research should focus on expanding training datasets and refining the model architecture to address these considerations, ensuring a more reliable and widely applicable tool for gender determination in forensic anthropology contexts.

Ethical considerations in our study, which involves the use of human bone images and data in forensic anthropology research, are fundamental to our research approach. Importantly, the data used in our study does not contain any private identity information of the patients, ensuring the ethical use of information in our research. As ethical standards evolve, ongoing oversight and a commitment to responsible engagement with human bone images and data remain integral to the ethical framework of our forensic anthropology research.

5. Conclusion

In conclusion, this paper presents a comprehensive exploration of the usability of bone parts, both individually and collectively, for sex determination. Leveraging the hybrid machine learning approach combining Artificial Neural Network (ANN) and Support Vector Machine (SVM), our study achieved a commendable accuracy of 83.33%, as evidenced by the Area Under Curve (AUC) metric. The pelvis and femur bones emerge as promising elements for accurate sex determination, showcasing their potential significance in forensic anthropology. Importantly, our findings highlight the impact of the chosen classification technique on accuracy, emphasizing the nuanced relationship between skeletal components and classification outcomes. The implication of this research extends to forensic practitioners, offering a reliable and accessible method for gender determination in cases where DNA analysis may be impractical. This study contributes not only to the advancement of forensic anthropology techniques but also underscores the broader implications for practical applications in real-world forensic investigations, making it a valuable and informative contribution to the field.

Acknowledgement

The authors are grateful to the Research Institute and Community Service (Institut Bisnis dan Teknologi Pelita Indonesia), Ministry of Education, Culture, Research and Technology Directorate General of Higher Education, and LLDIKTI Region X for the Applied Research Grant Scheme vote number 080/E5/PG.02.00.PL/2023, which have facilitated the success of this project.

References

Afrianty, I., Nasien, D., & Haron, H. (2022). Performance Analysis of Support Vector Machine in Sex Classification of The Sacrum Bone in Forensic Anthropology. *JURNAL TEKNIK INFORMATIKA*, *15*(1), 63–72. https://doi.org/10.15408/JTI.V15I1.25254

Afrianty, I., Nasien, D., Kadir, M. R. A., Haron, H., Azar, A. T., & Vaidyanathan, S. (2015). Back-Propagation neural network for gender determination in forensic anthropology. *Studies in Computational Intelligence*, *575*, 255–281. https://doi.org/10.1007/978-3-319-11017-2_11

Akhlaghi, M., Azizian, A., Sadeghian, M. H., Azizian, F., Shahabi, Z., Rafiee, S., & Mousavi, F. (2019). Collo-Diaphyseal Angle as an Optimal Anthropometric Criterion of Femur in Gender Determination. *International Journal of Medical Toxicology and Forensic Medicine*, *9*(2), 65–74. https://doi.org/10.32598/ijmtfm.v9i2.24986

Al-Boeridi, O. N., Syed Ahmad, S. M., & Koh, S. P. (2015). A scalable hybrid decision system (HDS) for Roman word recognition using ANN SVM: study case on Malay word recognition. *Neural Computing and Applications*, 26, 1505-1513.

Alfsdotter, C. (2021). Forensic archaeology and forensic anthropology within Swedish law enforcement: current state and suggestions for future developments. *Forensic Science International: Reports*, *3*. https://doi.org/10.1016/j.fsir.2021.100178

Anil K, J., & Jianchang, M. (1996). *Artificial Neural Networks: A Tutorial*. http://csc.lsu.edu/~jianhua/nn.pdf

Belousov, A. I., Verzakov, S. A., & Von Frese, J. (2002). Applicational aspects of support vector machines. *Journal of Chemometrics*, *16*(8–10), 482–489. https://doi.org/10.1002/CEM.744

Benz, P., Zhang, C., Karjauv, A., & Kweon, I. S. (2021). Robustness may be at odds with fairness: An empirical study on class-wise accuracy. *NeurIPS 2020 Workshop on Pre-Registration in Machine Learning*, 325–342.

Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: A critical evaluation. *BMC Medical Informatics and Decision Making*, *16*(3), 197–208. https://doi.org/10.1186/S12911-016-0318-Z/TABLES/5

Cattaneo, C. (2007). Forensic anthropology: developments of a classical discipline in the new millennium. *Forensic Science International*, *165*(2–3), 185–193. https://doi.org/10.1016/J.FORSCIINT.2006.05.018

Christensen, A. M., Leslie, W. D., & Baim, S. (2014). Ancestral differences in femoral neck axis length: possible implications for forensic anthropological analyses. *Forensic Science International*, *236*, 193.e1-193.e4. https://doi.org/10.1016/J.FORSCIINT.2013.12.027

Christensen, A. M., Passalacqua, N. V., Bartelink, E. J., Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (2014). Forensic Anthropology: Current Methods and Practice. In *Forensic Anthropology* (1st Edition). Academic Press.

Curate, F., Coelho, J., Gonçalves, D., Coelho, C., Ferreira, M. T., Navega, D., & Cunha, E. (2016). A method for sex estimation using the proximal femur. *Forensic Science International*, *266*, 579.e1-579.e7. https://doi.org/10.1016/J.FORSCIINT.2016.06.011

Darmawan, M. F. Bin, Osman, M. Z., & Nasien, D. (2021). Sex Estimation Model for Asian based on Random Forest Using Length of Left-Hand Bone. *2021 2nd International Conference on Artificial Intelligence and Data Sciences, AiDAS 2021*. https://doi.org/10.1109/AIDAS53897.2021.9574387

Dogan, N., & Tanrikulu, Z. (2013). A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness. *Information Technology and Management,* 14(2), 105–124. https://doi.org/10.1007/s10799-012-0135-8

Erlin, Marlim, Y. N., Junadhi, Suryati, L., & Agustina, N. (2022a). Early Detection of Diabetes Using Machine Learning with Logistic Regression Algorithm. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *11*(2), 88–96.

Erlin, Marlim, Y. N., Junadhi, Suryati, L., & Agustina, N. (2022b). Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *11*(2), 88–96. https://doi.org/10.22146/JNTETI.V11I2.3586

Fan, A. Z., & Koutris, P. (2022). Certifiable robustness for nearest neighbor classifiers. ArXiv Preprint ArXiv:2201.04770

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *KDD*.

Franchini, G., Ruggiero, V., Porta, F., Zanni, L., Franchini, G., Ruggiero, V., Porta, F., & Zanni, L. (2023). Neural architecture search via standard machine learning methodologies. *Mathematics in Engineering 2023 1:1*, *5*(1), 1–21. https://doi.org/10.3934/MINE.2023012

Gama, I., Navega, D., & Cunha, E. (2015). Sex estimation using the second cervical vertebra: a morphometric analysis in a documented Portuguese skeletal sample. *International Journal of Legal Medicine*, *129*(2), 365–372. https://doi.org/10.1007/S00414-014-1083-0

Gorodetsky, V., & Samoylov, V. (2008). Feature Extraction for Machine Learning : Logic – Probabilistic Approach. *The Fourth Workshop on Feature Selection in Data Mining*, 55–65.

Gorunescu, F. (2011). Data mining: Concepts, models and techniques. *Intelligent Systems Reference Library*, *12*. https://doi.org/10.1007/978-3-642-19721-5/COVER

Hairuddin, N. L., Yusuf, L. M., Othman, M. S., & Nasien, D. (2021). Gender classification using a pso-based feature selection and optimised bpnn in forensic anthropology. *International Journal of Computer Aided Engineering and Technology*, *15*(2–3), 232–242. https://doi.org/10.1504/IJCAET.2021.117133

Holland, T. D. (1991). Sex assessment using the proximal tibia. *American Journal of Physical Anthropology*, *85*(2), 221–227. https://doi.org/10.1002/AJPA.1330850210

Korkmaz Tan, R., & Bora, Ş. (2017). Parameter tuning algorithms in modeling and simulation. *International Journal of Engineering Science and Application*, 1(2), 58–66.

Kranioti, E. F., Bastir, M., Sánchez-Meseguer, A., & Rosas, A. (2009). A geometric-morphometric study of the Cretan humerus for sex identification. *Forensic Science International,* *189*(1–3), 111.e1-111.e8. https://doi.org/10.1016/J.FORSCIINT.2009.04.013

Lian, J., Freeman, L., Hong, Y., & Deng, X. (2021). Robustness with respect to class imbalance in artificial intelligence classification algorithms. *Journal of Quality Technology*, 53(5), 505–525.

Lee, S., Lee, S., Lim, J., & Lee, S. (2013). Defect Diagnostics of Power Plant Gas Turbine Using Hybrid SVM-ANN Method. *ASME 2012 Gas Turbine India Conference, GTINDIA 2012*, 725–732. https://doi.org/10.1115/GTINDIA2012-9564

Macaluso, P. J., & Lucena, J. (2014a). Estimation of sex from sternal dimensions derived from chest plate radiographs in contemporary Spaniards. *International Journal of Legal Medicine*, *128*(2), 389–395. https://doi.org/10.1007/S00414-013-0910-Z/METRICS

Macaluso, P. J., & Lucena, J. (2014b). Estimation of sex from sternal dimensions derived from chest plate radiographs in contemporary Spaniards. *International Journal of Legal Medicine*, *128*(2), 389–395. https://doi.org/10.1007/S00414-013-0910-Z/METRICS

Mays, S. (1992). Taphonomic factors in a human skeletal assemblage. *Circaea*, *9*(2), 54–58.

Meeusen, R. A., Christensen, A. M., & Hefner, J. T. (2015). The Use of Femoral Neck Axis Length to Estimate Sex and Ancestry. *Journal of Forensic Sciences*, *60*(5), 1300–1304. https://doi.org/10.1111/1556-4029.12820

Mohamad, M., Selamat, A., Krejcar, O., Fujita, H., & Wu, T. (2020). An analysis on new hybrid parameter selection model performance over big data set. *Knowledge-Based Systems*, 192, 105441. https://doi.org/10.1016/j.knosys.2019.105441

Nasien, D., Adiya, M. H., Afrianty, I., Ali, N. A., Samah, A. A., & Rahayu, Y. (2021, September). Determination of Sex and Race in Forensic Anthropology: A Comparison of Artificial Neural Network and Support Vector Machine. *In 2021 4th International Conference of Computer and Informatics Engineering* (IC2IE) (pp. 51-55). IEEE.

Navega, D., Vicente, R., Vieira, D. N., Ross, A. H., & Cunha, E. (2015). Sex estimation from the tarsal bones in a Portuguese sample: a machine learning approach. *International Journal of Legal Medicine*, *129*(3), 651–659. https://doi.org/10.1007/S00414-014-1070-5

Papaioannou, V. A., Kranioti, E. F., Joveneaux, P., Nathena, D., & Michalodimitrakis, M. (2012). Sexual dimorphism of the scapula and the clavicle in a contemporary Greek population: applications in forensic identification. *Forensic Science International*, *217*(1–3), 231.e1-231.e7. https://doi.org/10.1016/J.FORSCIINT.2011.11.010

Paya, B. A., Esat, I. I., & Badi, M. N. M. (1997). Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mechanical Systems and Signal Processing*, *11*(5), 751–765. https://doi.org/10.1006/MSSP.1997.0090

Phenice, T. W. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, *30*(2), 297–301. https://doi.org/10.1002/AJPA.1330300214

Rabunal, J. R., & Dorado, J. (2005). Artificial Neural Networks in Applications. In *SciencesNew York*. IGI Global. http://www.amazon.com/dp/1591409020

Santosh, K. C., Pradeep, N., Chakrabarti, T., Chakrabarti, P., Elngar, A. A., Nami, M., ... & Akbar, M. A. (2022). Performance Evaluation of LIBSVM and MSVM in Human Age Estimation and Gender Identification from Digital Images of Femur bone.

Shamim, N., & Yogesh. (2021). Machine Learning Based Feature Extraction of an Image: A Review. *Proceedings of International Conference on Machine Intelligence and Data Science Applications: MIDAS 2020*, 369–383.

Shylaja, S., & Muralidharan, R. (2019). Hybrid SVM-ANN Classifier is used for Heart Disease Prediction System. *INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH*, *07*(3). www.ijedr.org

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524. https://doi.org/10.1016/j.asoc.2019.105524

Spradley, M. K., & Jantz, R. L. (2011). Sex Estimation in Forensic Anthropology: Skull Versus Postcranial Elements. *Journal of Forensic Sciences*, *56*(2), 289–296. https://doi.org/10.1111/J.155

Stewart, T. D. (Thomas D. (1979). *Essentials of forensic anthropology, especially as developed in the United States*. Thomas.

Stojanowski, C. M. , & S. R. M. (1999). A reevaluation of the sex prediction accuracy of the minimum supero-inferior femoral neck diameter for modern individuals. *Journal of Forensic Sciences*, *44*(6), 1215–1218.

Suarez-Alvarez, M. M., Pham, D. T., Prostov, M. Y., & Prostov, Y.I. (2012). Statistical approach to normalization of feature vectors and clustering of mixed datasets. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2145), 2630–2651. https://doi.org/10.1098/rspa.2011.0704

Thomas, R. M., Parks, C. L., & Richard, A. H. (2017). Accuracy Rates of Ancestry Estimation by Forensic Anthropologists Using Identified Forensic Cases. *Journal of Forensic Sciences*, 62(4), 971–974. https://doi.org/10.1111/1556-4029.13361

White, T. D., Black, M. T., Folkens, P. A., White, T. D., Black, M. T., & Folkens, P. A. (2012). Human Osteology. In *Human Osteology*. Academic Press. http://www.sciencedirect.com:5070/book/9780123741349/human-osteology

Yaşar Işcan, M., & Olivera, H. E. S. (2000). Forensic anthropology in Latin America. *Forensic Science International*, *109*(1), 15–30. https://doi.org/10.1016/S0379-0738(99)00213-3