SINTECHCOM: Science, Technology and Communication Journal
Vol. 7, No. 1, October 2026, pp. 83-96, DOI: 10.59190/stc.v7i1.419

Comparative analysis of feature selection methods for support vector machine-based sex classification using craniometric data

Iis Afrianty[1], Fitri Wulandari[2,*], Nabilla Alya Firana[1]
[1]Department of Informatics Engineering, UIN Sultan Syarif Kasim Riau, Pekanbaru 28293, Indonesia
[2]Department of Informatics Engineering, UIN Sunan Kalijaga Yogyakarta, Yogyakarta 55281, Indonesia

ABSTRACT

Sex classification is a fundamental stage in forensic anthropology that can be supported by machine learning to achieve objective biological profiling. To address the high dimensionality and inter-trait multicollinearity of craniometric measurements, this study introduces a comparative feature-optimization framework that evaluates two distinct filtering paradigms entropy-based information gain (IG) and redundancy-aware correlation-based feature selection (CFS) integrated with a support vector machine (SVM) classifier against a full-feature baseline model. Experiments were conducted using the William W. Howells craniometric dataset, comprising 2,524 human crania (1,368 males and 1,156 females) characterized by 82 osteometric predictor variables after removing administrative identifiers. Model performance on a 70:30 training-testing split was evaluated using accuracy, precision, recall, F1-score, and the Matthews correlation coefficient (MCC). Experimental results show that the baseline SVM using all 82 features achieved an accuracy of 81.10% and an F1-score of 0.8110. Integrating feature selection substantially improved classification performance while compressing data dimensionality. The IG–SVM model achieved the highest accuracy of 86.73% (F1-score = 0.8674, MCC = 0.7348) using 20 selected features (a 75.60% feature reduction). Meanwhile, the CFS–SVM model achieved a virtually identical accuracy of 86.69% (F1-score = 0.8670, MCC = 0.7335) while requiring only 10 features (an 87.80% feature reduction). Crucially, these findings demonstrate that eliminating 75.60% to 87.80% of the original craniometric features not only enhances SVM classification accuracy by more than 5.5 percentage points by removing noisy and collinear variables, but also substantially reduces osteometric measurement complexity. Consequently, the parsimonious 10-feature CFS–SVM configuration provides an efficient and interpretable framework for computer-assisted sex estimation in forensic anthropology.

ARTICLE INFO

Article history:

Received Aug 21, 2026
Revised Sep 26, 2026
Accepted Oct 3, 2026
Published Oct 31, 2026

Keywords:

Craniometric Data
Feature Selection
Information Gain
Sex Classification
Support Vector Machine

Copyright (c) 2026:

Iis Afrianty, Fitri Wulandari, Nabilla Alya Firana

License:

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

***Corresponding Author**
E-mail address: fitri.wulandari@uin-suka.ac.id

1. INTRODUCTION

Forensic anthropology focuses on analyzing human skeletal remains to establish a biological profile comprising sex, age at death, stature, and population affinity for unidentified individuals [1-3]. Among these parameters, sex estimation is a pivotal initial step because subsequent estimations of age and stature are inherently sex-dependent [2,4]. Accurate sex assessment assists medicolegal investigations involving disaster victims, unidentified remains, or taphonomically damaged skeletons [5]. Next to the pelvis, the human cranium is the most widely utilized anatomical indicator due to its physical durability and pronounced sexually dimorphic traits [4,6-8]. Skeletal sex assessment is traditionally performed using morphological (visual) or osteometric (linear caliper or 3D landmark) approaches [9].

While morphological scoring is susceptible to observer subjectivity, conventional osteometric analyses can be hindered by inter-population cranial variation and bone fragmentation [2,10]. Moreover, molecular DNA profiling is frequently infeasible when biological tissues are severely degraded by burning or environmental exposure [11]. Consequently, machine learning algorithms offer a compelling computational alternative for objective and reproducible craniometric sex classification [9,10].

Several previous studies in forensic anthropology have integrated machine learning with cranial morphometrics, yet important methodological limitations persist regarding high-dimensional feature optimization. Toneva et al. [10] and Santos Mota et al. [3] demonstrated the efficacy of support vector machines (SVM) and artificial neural networks (ANN) for cranial sex estimation. On the 82-feature Howells craniometric dataset, recent studies have either applied SVM and ensemble classifiers across the full feature space without dimensionality reduction [12] or examined a single filter method paired with alternative classifiers, such as information gain (IG) with random forest [13] and correlation-based feature selection (CFS) with backpropagation neural networks [5]. However, prior studies have not systematically compared how distinct filter mechanisms specifically univariate entropy reduction (IG) versus correlation-based redundancy elimination (CFS) affect the margin optimization, classification accuracy, and feature compactness of SVM on high-dimensional craniometric data.

Craniometric datasets comprising dozens of measurement variables inevitably contain irrelevant or mutually redundant attributes that can induce overfitting and degrade classifier stability [14-16]. Not all cranial chords, subtenses, and angles contribute meaningfully to sex differentiation [14]. Filter-based feature selection techniques such as IG and CFS have proven effective in pruning uninformative attributes while retaining the most discriminative predictors [5,17,18]. Beyond reducing computational complexity, feature reduction enhances the anatomical interpretability of classification models in forensic anthropology [19].

The benchmark craniometric dataset compiled by Howells (1973, 1989, and 1995), comprising 2,524 human crania across 82 osteometric parameters, exemplifies this high-dimensional challenge. Because many cranial chords, subtenses, and radii measure overlapping anatomical regions, the dataset exhibits substantial multicollinearity and measurement noise that can degrade the margin optimization of SVM classifiers [20]. Despite the widespread adoption of SVM in osteometric sex assessment [9,10,21,22], a clear research gap remains: prior investigations have rarely compared how entropy-based ranking versus correlation-based redundancy reduction affects the trade-off between feature subset compactness and SVM classification accuracy. Addressing this gap is critical because forensic practitioners require models that achieve high diagnostic accuracy using a minimal number of cranial landmarks rather than requiring over 80 caliper measurements.

The specific novelty of this study lies in systematically contrasting two fundamentally different filter mechanisms univariate entropy reduction via IG and linear redundancy elimination via CFS to optimize maximum-margin SVM decision boundaries on high-dimensional craniometric data. Rather than focusing solely on classification accuracy over an exhaustive set of 82 cranial measurements, this research establishes the trade-off between extreme dimensionality reduction (retaining only 10 to 20 key anatomical traits) and classification stability. By identifying a compact, non-redundant subset of sexually dimorphic cranial parameters that outperforms the full-feature baseline, this study provides an efficient and interpretable computational framework for computer-assisted sex estimation in forensic anthropology [2,9,23].

2. RESEARCH METHODS

This study evaluates two filter-based feature selection methods IG and CFS integrated with an SVM classifier. Filter methods were selected over wrapper or embedded approaches due to their computational efficiency and independence from specific classifier induction biases [24]. Specifically, IG and CFS were chosen to compare two distinct mathematical selection criteria on morphometric data: IG is an information-theoretic univariate filter that ranks attributes individually by their capacity to reduce class entropy, whereas CFS evaluates linear feature-class correlation while suppressing inter-feature redundancy caused by collinear cranial dimensions [24].

2.1. Data Collection

Secondary osteometric data were obtained from the William W. Howells craniometric dataset (https://web.utk.edu/~auerbach/HOWL.htm), which compiles caliper-based linear lengths, breadths,

chords, subtenses, radii, and cranial angles collected by Dr. William W. Howells (1973, 1989, 1995). Although the highest recorded specimen identifier (ID) is 3,190 due to historical cataloging gaps in the archival source, the complete dataset contains N = 2,524 adult human crania, consisting of 1,368 male (54.20%) and 1,156 female (45.80%) specimens across 85 initial attributes and one target class (sex) [5]. Prior to feature selection and classification, three preprocessing steps were executed in Python (Google Colaboratory) [25]. First, non-craniometric attributes specifically the record ID, numerical population code (PopNum), and population name (Population) were removed so that classification relied exclusively on osteometric morphology, leaving 82 continuous craniometric predictor features and one target variable (Sex). Second, data integrity checks confirmed the absence of duplicate records or missing values across the 2,524 specimens. Third, the categorical target variable (sex) was binary-encoded as 1 for Male and 0 for Female, and predictor variables were standardized prior to SVM modeling. Representative feature definitions and sample records are presented in Tables 1 and Table 2.

Table 1. Skull bone parameter features.

Code	Skull bone parameter feature
GOL	Glabello-occipital length
NOL	Nasio-occipital length
BNL	Basion-nasion length
BBH	Basion-bregma height
XCB	Maximum cranial breadth
XFB	Maximum frontal breadth
ZYB	Bizygomatic breadth
…	…
TBA	Thiobarbituric acid

Table 2. Skull bone measurement parameters in mm.

ID	Sex	PopNum	Population	GOL	NOL	BNL	BBH	XCB	XFB	ZYB	⋯	TBA
1	M	1	NORSE	189	185	100	135	143	120	133	⋯	0
2	M	1	NORSE	182	178	102	139	145	120	137	⋯	0
3	M	1	NORSE	191	187	102	123	140	114	134	⋯	0
…	…	…	…	…	…	…	…	…	…	…	…	…
3188	F	24	ANDAMAN	160	160	89	121	129	106	117	⋯	156
3189	F	24	ANDAMAN	172	170	92	118	137	110	114	⋯	167
3190	F	24	ANDAMAN	153	153	88	116	130	105	115	⋯	158

Based on Table 2, the highest recorded ID reaches 3,190, while the number of samples used is 2,524. This discrepancy is due to the loss of some samples from the original data source, so not all IDs in that range have available data. Thus, the actual number of samples is lower than the highest recorded ID [5].

2.2. Data Preprocessing

The data preprocessing stage is essential because it converts raw data into a form suitable for the model, improving performance and ensuring accurate results [25]. Data preprocessing includes the following steps:

a. Data selection: This step aims to select relevant features to avoid unnecessary redundancy or duplication. The selected data is then stored in a file separate from the operational database [26]. In this study, the ID and Population features were removed from the dataset because ID merely provides a unique identifier for each record, while population merely restates the population name already coded in the PopNum feature. Relevant-feature selection was subsequently performed using IG and CFS.
b. Data cleaning: Data cleaning addresses missing values and data noise while ensuring consistency and relevance. In this study, checks were performed for duplicate and missing data; any such records were removed to ensure dataset quality.

c. Data transformation: This step converts the data into a format suitable for the model, which requires numeric input. Transformation was applied to the Sex feature, where Male was converted to 1 and Female was converted to 0.

2.3. Correlation-based Feature Selection (CFS)

CFS uses a correlation matrix to assess the relationship between features and the target class while ensuring that selected features are not highly correlated with one another. In this study, the correlation between each feature and the target was computed using the Pearson correlation, and features were selected via correlation-based attribute evaluation. The CFS feature-selection procedure is as follows [5]:

a. Prepare data with features (X) and target (Y).
b. Compute the correlation matrix to measure the relationship between each feature and the target, as well as between each pair of features, using the Pearson correlation coefficient in Equation (1) [27].

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (1)$$

where r is the correlation coefficient, x_i is the value of the i-th feature, y_i is the value of the i-th target, $\bar{x}$ is the mean of the feature, and $\bar{y}$ is the mean of the target.

c. Rank the features according to their correlation coefficient with the target.
d. Set a threshold value to determine relevant features.
e. Select features via correlation attribute evaluation: features with a correlation higher than the threshold are considered relevant to the target.

2.4. Feature Selection: Information Gain (IG)

Feature selection techniques are widely applied across domains such as text classification, genomic analysis, intrusion detection, and bioinformatics. IG is one of the most commonly used feature-selection methods; it reduces the number of irrelevant features and the dimensionality of the data to be analyzed. The calculation steps are as follows [18]:

a. Compute the entropy value. Entropy measures class uncertainty based on the probability of a given event or feature, per Equation (2).

$$Entropy(S) = \sum_i^n = 1 - pi\ log2\ pi \quad (2)$$

b. Compute the Information Gain using Equation (3).

$$Gain\ (S, A) = Entropy\ (S) - \sum_{values\ (A)} \frac{|Sv|}{|S|}\ Entropy\ (Sv) \quad (3)$$

where:

$Gain\ (S, A)$ = Information gain for feature A
$Entropy\ (S)$ = Total entropy value for all criteria of a feature
A = Feature
$Values(A)$ = The set of possible values for A
$|Sv|$ = Number of samples with value v
$|S|$ = Total number of data samples
$Entropy(Sv)$ = Entropy value of the samples with a particular value

2.5. Support Vector Machine (SVM)

SVM is one of the most widely used classification algorithms in machine learning research, particularly for two-class classification problems. Its main principle is to find a hyperplane, or linear separating plane, that maximizes the distance between classes referred to as the margin, enabling the model to distinguish data from two classes with a high level of generalization [8]. Commonly used kernel functions include those shown in Equation (4), Equation (5), and Equation (6) below:

$$K(x_i, x) = X_i^T x \tag{4}$$

$$K(x, y) = \exp\{-\frac{|x-y|^2}{2\gamma^2}\} \tag{5}$$

$$K(x_i, x) = (\gamma(x_i^T\ x) + r) \tag{6}$$

2.6. Model Evaluation and Testing

This study tested feature selection methods to optimize SVM classification for determining sex from craniometric measurements. Evaluation was carried out using a confusion matrix, a table presenting the total number of correctly and incorrectly categorized data points, which facilitates assessment of a classification model's accuracy [28]. The confusion matrix consists of four components: true positive (TP), positively labeled data correctly classified as positive; true negative (TN), negatively labeled data correctly classified as negative; false positive (FP), positively labeled data incorrectly classified as negative; and false negative (FN), negatively labeled data incorrectly classified as positive. The confusion matrix representation is shown in Figure 1.

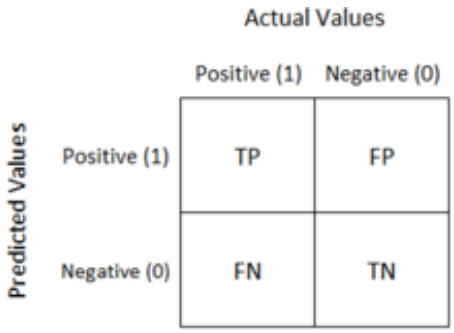


Figure 1. Confusion matrix.

From the confusion matrix, several evaluation metrics can be calculated as follows [29]:

a. Accuracy: The degree to which predictions match actual data, i.e., the ratio of correctly categorized data to total test data, calculated using Equation (7).

$$Accuracy\ = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

b. Precision: The accuracy of positive predictions, computed as TP / (TP + FP), Equation (8).

$$Precision\ = \frac{TP}{TP+FP} \tag{8}$$

c. Recall: The proportion of actual positive data correctly identified, computed as TP / (TP + FN), Equation (9).

$$Recall\ = \frac{TP}{TP+FN} \tag{9}$$

d. F1-score: The harmonic balance between precision and recall, particularly useful under class imbalance, Equation (10).

$$F1\ Score\ = \frac{2 \times (Precision \times Recall)}{Presisi + Recall} \tag{10}$$

3. RESULTS AND DISCUSSIONS

This study evaluates the effect of applying two feature selection methods, IG and CFS, on the performance of the SVM classification algorithm for sex identification using human craniometric data.

The analysis assesses the extent to which removing irrelevant features improves accuracy and efficiency compared with using all features.

Table 3. Correlation coefficients for all features.

No	Feature	Correlation coefficient	No	Feature	Correlation coefficient
1	ZYB	0.611964	42	NLB	0.244976
2	JUB	0.592574	43	PAF	0.234797
3	MDH	0.546812	44	FRA	0.229258
4	GOL	0.510446	45	DKB	0.225336
5	FMB	0.505368	46	OCC	0.221918
6	ZMB	0.492463	47	NDS	0.221706
7	MDB	0.490153	48	SSS	0.204639
8	NOL	0.482993	49	SIS	0.202681
9	XML	0.477645	50	PAS	0.192977
10	MAB	0.472692	51	STB	0.190286
11	EKB	0.471429	52	NAS	0.187965
12	NAR	0.467284	53	OCF	0.171815
13	SOS	0.464924	54	BAR	0.135124
14	AUB	0.456465	55	OCS	0.135124
15	BNL	0.454710	56	OBH	0.129992
16	BBH	0.440208	57	BAA	0.124508
17	GLS	0.438358	58	BRR	0.113988
18	FRC	0.432319	59	OSR	0.112126
19	SSR	0.428763	60	LAR	0.105086
20	VRR	0.422895	61	BSA	0.084295
21	AVR	0.413644	62	SLA	0.083354
22	DKR	0.410510	63	RPA	0.078590
23	NLH	0.407801	64	SBA	0.076313
24	NPH	0.404381	65	RFA	0.074606
25	PRR	0.396714	66	BRA	0.074582
26	ASB	0.385830	67	TBA	0.072723
27	FMR	0.383205	68	ROA	0.067247
28	FRF	0.381618	69	DKA	0.048765
29	WCB	0.381452	70	DKS	0.045034
30	ZOR	0.373927	71	PRA	0.044715
31	EKR	0.371899	72	NBA	0.005343
32	OBB	0.366074	73	SSA	0.004412
33	PAC	0.361261	74	FRS	-0.004782
34	ZMR	0.356341	75	WNB	-0.006584
35	IML	0.347209	76	PAA	-0.028450
36	XCB	0.340279	77	OCA	-0.035412
37	FOL	0.337494	78	NFA	-0.039471
38	BPL	0.336499	79	BBA	-0.068878
39	XFB	0.336157	80	NDA	-0.077244
40	WMH	0.317293	81	NAA	-0.124315
41	MLS	0.284454	82	SIA	-0.223621

3.1. Data Collection

Feature selection was performed by computing Pearson's correlation coefficient between each of the 82 craniometric predictors and the target sex label (see Table 3). Based on the ranked correlation coefficients in Table 3, progressive thresholding was analyzed to determine an optimal feature subset that maximizes target relevance while minimizing inter-feature redundancy. Applying permissive thresholds of $|r| \geq 0.01$ and $|r| \geq 0.10$ retained 72 and 60 features, respectively; however, these

broad subsets still included dozens of weakly dimorphic cranial angles and mutually collinear chord measurements that degrade SVM margin optimization. Conversely, examining the upper tail of the correlation distribution revealed a natural cluster of 10 primary osteometric dimensions exceeding a strict threshold of $|r| > 0.47$ (ZYB [0.6120], JUB [0.5926], MDH [0.5468], GOL [0.5104], FMB [0.5054], ZMB [0.4925], MDB [0.4902], NOL [0.4830], XML [0.4776], and MAB [0.4727]). Selecting this top-10 subset eliminated 72 lower-relevance and redundant variables (an 87.80% dimensionality reduction) while preserving the core craniofacial breadth, mastoid, and vault length dimensions for CFS–SVM classification [24].

Table 4. IG values for all features.

Feature	Information gain	Feature	Information gain
ASB	0.127848271	NLB	0.048992675
AUB	0.184492768	NLH	0.14379941
AVR	0.143916119	NOL	0.205257669
BAA	0.015448905	NPH	0.145542502
BAR	0.037480992	OBB	0.107799445
BBA	0.008719277	OBH	0.019757622
BBH	0.170530935	OCA	0.012147055
BNL	0.176754638	OCC	0.045435106
BPL	0.1009411	OCF	0.033228754
BRA	0.010451491	OCS	0.020676415
BRR	0.127734752	OSR	0.052135431
BSA	0.027382948	PAA	0.003205368
DKA	0.019172631	PAC	0.112220115
DKB	0.044552043	PAF	0.055527797
DKR	0.143044725	PAS	0.032198941
DKS	0.010468085	PRA	0.010856414
EKB	0.190466594	PRR	0.13582375
EKR	0.113758117	RFA	0.008085378
FMB	0.223080269	ROA	0.019510076
FMR	0.123460568	RPA	0.012758596
FOL	0.091566523	SBA	0.010726108
FRA	0.047850312	SIA	0.078313929
FRC	0.160686958	SIS	0.049989087
FRF	0.121628305	SLA	0.012496673
FRS	0.011918792	SOS	0.179671486
GLS	0.158686562	SSA	0.011866532
GOL	0.232582141	SSR	0.157058959
IML	0.098739429	SSS	0.040835354
JUB	0.322660278	STB	0.044645066
LAR	0.068972508	TBA	0.018431562
MAB	0.18658099	VRR	0.154433423
MDB	0.207877794	WCB	0.122412331
MDH	0.264109654	WMH	0.081576882
MLS	0.063742844	WNB	0.048661034
NAA	0.016873142	XCB	0.10999366
NAR	0.189654618	XFB	0.108968846
NAS	0.034223996	XML	0.193243595
NBA	0.01200173	ZMB	0.212248253
NDA	0.035569887	ZMR	0.110233769
NDS	0.043294906	ZOR	0.115164838
NFA	0.01125423	ZYB	0.354120049

3.2. Data Preprocessing

IG feature selection was initiated by computing the baseline Shannon entropy of the target variable (sex) using Equation (2), followed by the feature-specific expected information gain via Equation (3). As detailed in Table 4, the top 20 features exhibiting the highest IG (IG $>$ 0.1544, representing a 75.60% reduction from the 82 initial measurements) are ZYB (0.3541), JUB (0.3227), MDH (0.2641), GOL (0.2326), FMB (0.2231), ZMB (0.2122), MDB (0.2079), NOL (0.2053), XML (0.1932), EKB (0.1905), NAR (0.1897), MAB (0.1866), AUB (0.1845), SOS (0.1797), BNL (0.1768), BBH (0.1705), FRC (0.1607), GLS (0.1587), SSR (0.1571), and VRR (0.1544). Comparing this 20-feature IG subset with the 10-feature CFS subset demonstrates that reducing the measurement protocol from 82 variables to 10–20 traits retains the most diagnostically robust regions of the human cranium specifically the zygomatic arch and upper facial breadth, the mastoid process, and primary neurocranial length while eliminating 62 to 72 secondary subtenses, chords, and angles that are prone to landmark ambiguity and postmortem damage [4].

a. Data selection: The ID, PopNum, and population features were removed. The result is shown in Table 5, leaving 82 features with 1 class label.

Table 5. Result of data selection.

Sex	PopNum	GOL	NOL	BNL	BBH	XCB	XFB	ZYB	⋯	TBA
M	1	189	185	100	135	143	120	133	⋯	0
M	1	182	178	102	139	145	120	137	⋯	0
M	1	191	187	102	123	140	114	134	⋯	0
…	…	…	…	…	…	…	…	…	…	…
F	24	160	160	89	121	129	106	117	⋯	156
F	24	172	170	92	118	137	110	114	⋯	167
F	24	153	153	88	116	130	105	115	⋯	158

b. Data transformation: The sex feature was transformed, with male converted to 1 and female converted to 0, as shown in Table 6.

Table 6. Result of data transformation on the sex feature.

Sex	PopNum	GOL	NOL	BNL	BBH	XCB	XFB	ZYB	⋯	TBA
1	1	189	185	100	135	143	120	133	⋯	0
1	1	182	178	102	139	145	120	137	⋯	0
1	1	191	187	102	123	140	114	134	⋯	0
…	…	…	…	…	…	…	…	…	…	…
0	24	160	160	89	121	129	106	117	⋯	156
0	24	172	170	92	118	137	110	114	⋯	167
0	24	153	153	88	116	130	105	115	⋯	158

3.3. Correlation-Based Feature Selection (CFS)

Feature selection was performed by computing the Pearson correlation coefficient between each feature and the target, as shown in Equation (1). The correlation-coefficient results, ranked by their association with the target, are shown in Table 3.

Feature selection used correlation-based attribute evaluation, with a predefined threshold determining the minimum strength of the relationship between each feature and the classification target. A higher threshold results in stricter selection, retaining only features with stronger correlations and excluding those with weaker relationships to the target. With a threshold of 0.01, 72 features were selected, while a threshold of 0.1 yielded only 60 features. Increasing the threshold reduces the number of selected features but increases their relevance to the classification target. The selected features were subsequently used as input to the SVM classifier to evaluate the effect of feature selection on model performance, particularly classification accuracy and computational efficiency.

3.4. Information Gain (IG)

IG feature selection begins by computing the entropy value of the sex feature using Equation (2). The resulting IG values are shown in Table 4.

3.5. Model Evaluation and Testing

Three experimental scenarios were evaluated using a 70:30 training-to-testing split: (1) Baseline SVM without feature selection (82 features), (2) IG–SVM (20 features), and (3) CFS–SVM (10 features). On the Howells dataset (N = 2,524, comprising 54.20% male and 45.80% female crania), the 70:30 partitioning scheme allocates 1,767 specimens to train the SVM separating hyperplane while reserving a substantial holdout set of 757 unseen specimens for independent validation. This large holdout sample size (n = 757) ensures stable out-of-sample error estimation and mitigates the high variance commonly observed in small-sample osteometric studies. Furthermore, to ensure that model reliability is not distorted by the slight male-to-female sample ratio (1,368 vs. 1,156), performance was assessed not only via accuracy and F1-score, but also via MCC, which evaluates all four confusion matrix quadrants symmetrically (Table 7).

Table 7. Test results.

Feature selection method	Classification algorithm	Number of selected features	Precision	Recall	F1-score	MCC	Accuracy (%)
No feature selection	SVM	82	0.811	0.811	0.811	-	81.10
Information gain	SVM	20	0.8689	0.8673	0.8674	0.7348	86.73
CFS	SVM	10	0.8681	0.8669	0.8670	0.7335	86.69

As shown in Table 7, integrating feature selection with SVM substantially improved classification performance compared with the 82-feature baseline model, which achieved an accuracy of 81.10% and an F1-score of 0.8110. From a theoretical perspective, although SVM is a maximum-margin classifier capable of handling high-dimensional inputs, its decision boundary depends directly on inner products or pairwise Euclidean distances across all input coordinates. Examination of Table 3 and Table 4 reveals that more than one-third of the 82 Howells craniometric variables particularly cranial angles and minor subtenses such as SSA (r = 0.0044, IG = 0.0119), FRS (r = -0.0048, IG = 0.0119), NBA (r = 0.0053, IG = 0.0120), and WNB (r = -0.0066, IG = 0.0487) exhibit virtually zero association with biological sex. When all 82 features are fed into the baseline SVM, these non-dimorphic attributes act as spherical noise in the kernel space, perturbing pairwise sample distances, increasing the number of margin violations (slack variables), and degrading out-of-sample generalization [20]. Applying IG–SVM elevated classification accuracy by 5.63 percentage points to 86.73% (F1-score = 0.8674, MCC = 0.7348) while compressing the feature space by 75.60% to 20 variables. By pruning 62 low-entropy variables, IG–SVM eliminates noisy coordinate axes and enables the SVM optimizer to construct a wider, more stable separating margin.

After feature selection using IG, accuracy increased to 86.73%, with an F1-score of 0.8674 and an MCC of 0.7348. IG selects features based on the highest information value related to the target variable, retaining the features with the greatest influence on classification. Removing irrelevant attributes improved the data distribution in the vector space, allowing SVM to form a more optimal separating hyperplane, while the reduced number of features (20) made model training more efficient without sacrificing stability.

Meanwhile, the CFS–SVM model achieved an accuracy of 86.69%, an F1-score of 0.8670, and an MCC of 0.7335 using only 10 selected features, representing an 87.80% reduction from the initial 82 features. A critical comparison between IG–SVM and CFS–SVM justifies why CFS–SVM is recommended as the most parsimonious configuration despite IG–SVM achieving a nominally higher accuracy by 0.04 percentage points (and a 0.0013 difference in MCC). Across the 30% test partition ($n = 757$ crania), a 0.04 percentage-point difference corresponds to fewer than one skull classification difference, indicating virtual statistical parity. Cross-referencing Table 3 and Table 4 reveals why this parity occurs: the top nine features ranked by Pearson correlation and IG are identical in exact order. The additional 10 variables retained by IG–SVM (ranks 11 – 20) exhibit moderate collinearity with

the primary facial breadth and cranial vault length dimensions already captured in the 10-feature subset. Therefore, under the principle of parsimony, while IG–SVM represents the highest-scoring model in absolute accuracy, CFS–SVM offers the optimal trade-off between predictive power and osteometric efficiency by halving the required physical measurements without meaningful loss of classification reliability [24,30].

Beyond computational efficiency in training time and memory footprint, reducing the required input from 82 craniometric variables to 10 – 20 key measurements provides substantial operational value in forensic anthropology practice. First, acquiring all 82 Howells craniometric dimensions which include intricate facial subtenses, radii, and cranial base angles is labor-intensive, requires specialized spreading, sliding, and coordinate calipers, and increases the likelihood of inter-observer measurement error [1,3]. Second, forensic skeletal remains recovered from crime scenes or mass disasters are frequently subjected to taphonomic degradation, burning, or postmortem fragmentation, rendering a complete 82-variable osteometric profile unattainable [1,2]. By isolating the 10 most informative dimensions in the CFS–SVM model, forensic practitioners can focus exclusively on structurally durable and highly dimorphic anatomical regions specifically facial breadth, the mastoid process, and vault length [4]. These findings also carry direct implications for automated forensic identification systems. In modern forensic workflows, osteometric data are increasingly acquired either through digital calipers linked to laboratory information systems or via automated landmark detection on 3D CT and surface scans [2,9]. Embedding the parsimonious 10-feature CFS–SVM model into a forensic decision-support tool allows rapid, objective sex screening while enabling forensic anthropologists to easily verify, interpret, and communicate the evidentiary basis of the automated prediction in medicolegal contexts.

Within the holdout test partition of the Howells dataset, the SVM classifier effectively utilized the reduced feature subsets from IG and CFS to achieve stable out-of-sample classification performance. Despite these promising results, several limitations of this study must be acknowledged. First, the experimental evaluation relied on a single public benchmark repository the Howells craniometric dataset. Because patterns of cranial sexual dimorphism vary across contemporary populations due to secular change, genetics, and environmental factors [2], high holdout accuracy within the Howells collection does not automatically guarantee cross-population generalization to modern forensic casework. Second, model evaluation was conducted using a single 70:30 train-test split rather than repeated stratified cross-validation or external validation on an independent contemporary skeletal collection. Future studies should therefore validate the 10-feature CFS–SVM and 20-feature IG–SVM models across independent, documented osteometric and CT-derived datasets from diverse geographic regions.

4. CONCLUSION

Based on the experimental evaluation on the Howells craniometric dataset (N = 2,524), integrating filter-based feature selection with SVM substantially improved sex classification performance while compressing data dimensionality. The baseline SVM model using all 82 craniometric features achieved an accuracy of 81.10% and an F1-score of 0.8110. Applying IG (IG–SVM) yielded the highest accuracy of 86.73% (F1-score = 0.8674, MCC = 0.7348) using 20 selected features, corresponding to a 5.63 percentage-point accuracy gain and a 75.60% reduction in features. Concurrently, the CFS (CFS–SVM) model achieved a virtually identical accuracy of 86.69% (F1-score = 0.8670, MCC = 0.7335) using only 10 selected features, representing a 5.59 percentage-point accuracy gain and an 87.80% reduction in feature dimensionality. Overall, both IG and CFS effectively isolated the most sexually dimorphic craniometric traits (ZYB, JUB, MDH, GOL, FMB, ZMB, MDB, NOL, XML, and MAB) while eliminating noisy and collinear measurements. While IG–SVM achieved the highest absolute accuracy, CFS–SVM is recommended as the most operationally efficient configuration because it achieves comparable performance with half as many features. In practical forensic classification systems, cutting the required osteometric input by 87.80% accelerates caliper or CT-based landmark acquisition, accommodates partially fragmented crania where secondary measurements are missing, and enables lightweight deployment within computer-assisted forensic decision-support tools. Future research should validate these reduced-feature models across independent contemporary osteometric collections using repeated stratified cross-validation, explore systematic hyperparameter optimization and hybrid filter-wrapper selection, and integrate the 10-feature CFS–SVM model into a standalone forensic decision-support application.

AUTHOR CONTRIBUTIONS

Iis Afrianty: Conceptualization, methodology, research planning, supervision, development of the comparative evaluation framework, interpretation of findings, and writing—review and editing. Fitri Wulandari: Methodological refinement, validation of feature selection procedures, assessment of classification performance, verification of analytical consistency, critical interpretation of results, and writing—review and editing. Nabilla Alya Firana: Data curation, preprocessing, software implementation, feature selection experiments, SVM model training and testing, calculation of evaluation metrics, formal analysis, preparation of tables and visualizations, and writing—original draft. All authors contributed to discussing the findings, evaluating the study's limitations, and revising the manuscript for intellectual content. All authors approved the final version and agreed to be accountable for the accuracy and integrity of the work.

FUNDING STATEMENT

This research received no external funding.

DECLARATION OF AI USE

The authors declare that no AI tools were used in preparing this manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data supporting this study are included in the article. No separate datasets are available.

ETHICAL APPROVAL

Ethical approval was unnecessary, as no humans or animals were involved.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the funder for its financial support of this research in 2026, which enabled the implementation and completion of the study. The authors also express their sincere appreciation to Universitas Islam Negeri Sultan Syarif Kasim Riau for its institutional support and for providing the laboratory facilities and academic environment that supported the research activities, data analysis, interpretation of results, and preparation of this manuscript. These contributions facilitated the evaluation of feature selection methods and SVM classification models and supported the dissemination of the research findings.

REFERENCES

[1] Knecht, S., Santos, F., Ardagna, Y., Alunni, V., Adalian, P., & Nogueira, L. (2023). Sex estimation from long bones: A machine learning approach. *International Journal of Legal Medicine*, **137**(6), 1887-1895.
DOI: https://doi.org/10.1007/s00414-023-03072-4

[2] Wang, X., Liu, G., Wu, Q., Zheng, Y., Song, F., & Li, Y. (2024). Sex estimation techniques based on skulls in forensic anthropology: A scoping review. *PLoS ONE*, **19**(12), 1-22.
DOI: https://doi.org/10.1371/journal.pone.0311762

[3] Santos Mota, M. J., Alves Vieira, A. C., Silva Lima, L., Sátiro, J. V. M., de Menezes Neto, C. M., Prado Paixão, P. L., Gonçalves Lopes, G. P., de Azevedo Setton, L. R., de Andrade, C. E., & Halti Cabral, R. (2025). Enhancing sex determination in forensic anthropology: A comparative analysis of cranial measurements using artificial neural network. *Forensic Science International: Reports*, **12**, 100422.
DOI: https://doi.org/10.1016/j.fsir.2025.100422

[4] Mello-Gentil, T. & Souza-Mello, V. (2022). Contributions of anatomy to forensic sex estimation: focus on head and neck bones. *Forensic Sciences Research*, **7**(1), 11-23.
DOI: https://doi.org/10.1080/20961790.2021.1889136

[5] Ma'rifah, L. A., Afrianty, I., Budianita, E., & Syafria, F. (2025). Klasifikasi tulang tengkorak berdasarkan jenis kelamin menggunakan correlation-based feature selection (CFS) dengan backpropagation neural network (BPNN). *Jurnal Informatika: Jurnal Pengembangan IT*, **10**(2), 333–347.
DOI: https://doi.org/10.30591/jpit.v10i2.8616

[6] 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.
DOI: https://doi.org/10.15408/jti.v15i1.25254

[7] Afrianty, I., Mhd. Kadarman, Elvia Budianita, & Fadhilah Syafria. (2024b). Klasifikasi tulang tengkorak manusia berdasarkan jenis kelamin menggunakan backpropagation pada antropologi forensik. *Jurnal CoSciTech (Computer Science and Information Technology)*, **5**(3), 619-625.
DOI: https://doi.org/10.37859/coscitech.v5i3.8235

[8] Rahayu, S. S., Afrianty, I., Budianita, E., & Syafria, F. (2024). Klasifikasi Tulang Tengkorak Berdasarkan Jenis Kelamin dalam Antropologi Forensik Menggunakan Metode Support Vector Machine. *INOVTEK Polbeng - Seri Informatika*, **9**(1).
DOI: https://doi.org/10.35314/isi.v9i1.4046

[9] Faisal, E. & Rogers, T. L. (2025). A review of the literature on the applications of machine learning in forensic anthropology. *Forensic Science International*, **376**, 112579.
DOI: https://doi.org/10.1016/j.forsciint.2025.112579

[10] Toneva, D., Nikolova, S., Agre, G., Zlatareva, D., Hadjidekov, V., & Lazarov, N. (2021). Machine learning approaches for sex estimation using cranial measurements. *International Journal of Legal Medicine*, **135**(3), 951-966.
DOI: https://doi.org/10.1007/s00414-020-02460-4

[11] Nasien, D., Adiya, M. H., Afrianty, I., Ali, N. A., Samah, A. A., & Rahayu, Y. (2021). Determination of sex and race in forensic anthropology: a comparison of artificial neural network and support vector machine. *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, 51-55.
DOI: https://doi.org/10.1109/IC2IE53219.2021.9649182

[12] Ramadhani, A., Afrianty, I., Budianita, E., & Gusti, S. K. (2025). Implementation of XGBoost ensemble and support vector machine for gender classification of skull bones. *Bulletin of Informatics and Data Science*, **4**(1), 34-41.
DOI: https://doi.org/10.61944/bids.v4i1.115

[13] Afrianty, I., Mhd. Kadarman, Elvia Budianita, & Fadhilah Syafria. (2024a). Klasifikasi Tulang Tengkorak Manusia Berdasarkan Jenis Kelamin Menggunakan Backpropagation Pada Antropologi Forensik. *Jurnal CoSciTech (Computer Science and Information Technology)*, **5**(3), 619-625.
DOI: https://doi.org/10.37859/coscitech.v5i3.8235

[14] Hairuddin, N. L., Mi Yusuf, L., & Othman, M. S. (2020). Gender classification on skeletal remains: Efficiency of metaheuristic algorithm method and optimized back propagation neural network. *Journal of Information and Communication Technology*, **19**(2), 251-277.
DOI: https://doi.org/10.32890/jict2020.19.2.5

[15] Harni, Y., Afrianty, I., Sanjaya, S., Abdillah, R., Yanto, F., & Syafria, F. (2023). Performance analysis of LVQ 1 using feature selection gain ratio for sex classification in forensic anthropology. *Building of Informatics, Technology and Science (BITS)*, **5**(1).
DOI: https://doi.org/10.47065/bits.v5i1.3625

[16] Chakraborty, P., Bandyopadhyay, A., Sahu, P. P., Burman, A., Mallik, S., Alsubaie, N., Abbas, M., Alqahtani, M. S., & Soufiene, B. O. (2024). Predicting stroke occurrences: a stacked ma-

chine learning approach with feature selection and data preprocessing. *BMC Bioinformatics*, **25**(1), 329.
DOI: https://doi.org/10.1186/s12859-024-05866-8

[17] Fitri, A., Afrianty, I., Budianita, E., & Gusti, and S. K. (2025). Implementation of feature selection information gain in support vector machine method for stroke disease classification. *Bulletin of Informatics and Data Science*, **4**(1), 22-33.
DOI: https://doi.org/10.61944/bids.v4i1.116

[18] Khair, N. T., Afrianty, I., Syafria, F., Budianita, E., & Gusti, S. K. (2025). Penerapan information gain untuk seleksi fitur pada klasifikasi jenis kelamin tulang tengkorak menggunakan backpropagation. *Bulletin of Computer Science Research*, **5**(4), 666-678.
DOI: https://doi.org/10.47065/bulletincsr.v5i4.637

[19] Morgado, J., Pereira, T., Silva, F., Freitas, C., Negrão, E., de Lima, B. F., da Silva, M. C., Madureira, A. J., Ramos, I., Hespanhol, V., Costa, J. L., Cunha, A., & Oliveira, H. P. (2021). Machine learning and feature selection methods for EGFR mutation status prediction in lung cancer. *Applied Sciences*, **11**(7), 3273.
DOI: https://doi.org/10.3390/app11073273

[20] Kalagotla, S. K., Basanta, T., & Devi, M. B. (2026). Correlation-SVM: A multicollinearity-aware feature selection framework for SVM-based medical diagnosis. *Recent Trends in Artificial Intelligence & Its Applications*, **5**(2).
DOI: https://doi.org/10.46610/RTAIA.2026.v05i02.001

[21] Abdumalikov, S., Kim, J., & Yoon, Y. (2024). Performance analysis and improvement of machine learning with various feature selection methods for EEG-based emotion classification. *Applied Sciences*, **14**(22), 10511.
DOI: https://doi.org/10.3390/app142210511

[22] Ullah, T., Ullah, S. I., Ullah, K., Ishaq, M., Khan, A., Ghadi, Y. Y., & Algarni, A. (2024). Machine learning-based cardiovascular disease detection using optimal feature selection. *IEEE Access*, **12**, 16431-16446.
DOI: https://doi.org/10.1109/ACCESS.2024.3359910

[23] Spiros, M., Kamnikar, K., & Plemons, A. (2025). Anthropology: Population affinity estimation. *Encyclopedia of Forensic and Legal Medicine*, 255-267.
DOI: https://doi.org/10.1016/B978-0-443-21441-7.00016-9

[24] Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. Doctoral dissertation, University of Waikato.
URL: https://researchcommons.waikato.ac.nz

[25] Anargya, M. A. N., Ghozi, W., & Rafrastara, F. A. (2025). Random Under Sampling for Performance Improvement in Attack Detection on Internet of Vehicles Using Machine Learning. *Jurnal Informatika: Jurnal Pengembangan IT*, **10**(1), 11-19.
DOI: https://doi.org/10.30591/jpit.v10i1.8034

[26] Irawan, Y. (2019). Penerapan data mining untuk evaluasi data penjualan menggunakan metode clustering dan algoritma hirarki divisive di perusahaan media world Pekanbaru. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, **4**(1), 13-20.
DOI: https://doi.org/10.20527/jtiulm.v4i1.34

[27] Burgund, D., Nikolovski, S., Galić, D., & Maravić, N. (2023). Pearson correlation in determination of quality of current transformers. *Sensors*, **23**(5).
DOI: https://doi.org/10.3390/s23052704

[28] Nurhidayat, R. & Dewi, K. E. (2023). Penerapan algoritma K-nearest neighbor dan fitur ekstraksi N-gram dalam analisis sentimen berbasis aspek. *Komputa : Jurnal Ilmiah Komputer dan Informatika*, **12**(1), 91-100.
DOI: https://doi.org/10.34010/komputa.v12i1.9458

[29] Puteri, A. N., Arizal, A., & Achmad, A. D. (2021). Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, **20**(2), 335-342.
DOI: https://doi.org/10.30812/matrik.v20i2.1183

[30] Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Huthwelker, J., Drees, P., Fröhlich, M., & Betz, U. (2021). General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait. *Computer Methods in Biomechanics and Biomedical Engineering*, **24**(3), 299-307.
DOI: https://doi.org/10.1080/10255842.2020.1828375