COMP4702 PROJECT REPORT

BIOMETRIC EMBEDDINGS FOR TABLE TENNIS PLAYER IDENTIFICATION

STUDENT: **MAX GADD**

STUDENT NUMBER: **46985431**

DATE: **30/05/2025**

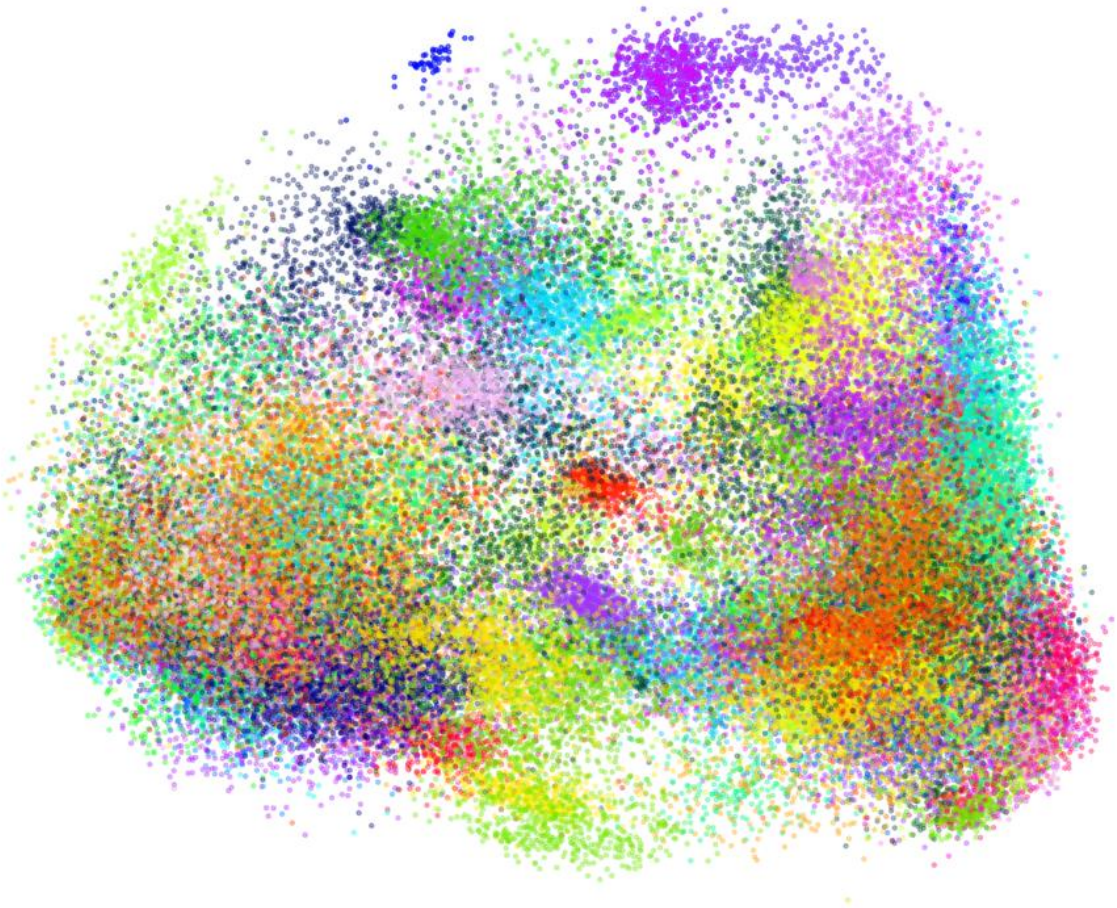

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This study evaluated techniques for identifying individual table-tennis players using 9-axis inertial sensor data embedded in racket grips. Initial feature engineering produced 36 summary statistics per stroke, enabling a k-nearest neighbours baseline that achieved around 80 percent accuracy but proved memory-heavy and skewed toward well-represented players. A Siamese triplet-loss network then generated 64-dimensional embeddings, though standalone performance plateaued near 60 percent. The final model built on this by fine-tuning the same embedding backbone with a linear classification head: a multilayer perceptron with batch normalisation, dropout and a 64-unit embedding layer, followed by a single softmax output layer over 93 player IDs. It was trained end-to-end using the AdamW optimiser, a One-Cycle learning-rate schedule, class-balanced sampling and weighted cross-entropy loss, yielding approximately 68 percent test accuracy and markedly better robustness to under-represented players. Future work may include the exploration of advanced metric-learning losses and per-exercise data balancing to further improve identification performance and ability to deploy.

BACKGROUND

This report uses the TTSWING dataset, which leverages wearable inertial measurement units (IMUs) to capture the detailed biomechanics of elite table-tennis strokes, enabling rich analysis of human motion in sports science and ubiquitous computing contexts [1]. Feature engineering in this domain combines time-domain, aggregate statistical, and frequency-domain summaries to distil high-frequency sensor streams into interpretable descriptors. Metric learning approaches, such as Siamese networks trained with triplet loss, further aim to structure these descriptors into discriminative embedding spaces that reflect stroke similarity and player identity.

DATA COLLECTION CONCEPTS

Wearable IMUs typically integrate tri-axial accelerometers, gyroscopes, and magnetometers to record multi-axis motion data, facilitating continuous monitoring of human movement patterns in real time. In sports biomechanics, custom embedding of a 9-axis IMU into a racket handle allows direct capture of stroke kinematics, avoiding occlusion and calibration issues common in camera-based systems [2].

EXPLORATORY DATA ANALYSIS CONCEPTS

Exploratory Data Analysis (EDA) of IMU features begins with univariate assessments such as histograms and density plots which are used to detect sensor faults, distribution skew, and tail behaviour [3]. Outlier handling in this context may involve clipping extreme values, robust down-weighting based on Huber loss, or modality-specific thresholding to maintain focus on representative stroke patterns.

CLASS IMBALANCE CONCEPTS

Imbalanced class distributions occur when players contribute uneven stroke counts, and this can bias classifiers toward majority classes. Inverse-frequency weighting applies class-specific loss multipliers proportional to the reciprocal of stroke counts, equalising training importance across under- and over-represented players [4].

BASELINE MODELLING CONCEPTS

Non-parametric methods like k-nearest neighbours (k-NN) classify new strokes based on proximity in feature space, offering intuitive baselines for IMU-based HAR without intensive parameter tuning [1]. k-NN performance inherently benefits from larger, more consistent class samples, reflecting the need for balanced data or weighting strategies to prevent majority-class dominance [5].

METRIC LEARNING CONCEPTS

Siamese networks trained with triplet loss optimise embeddings by minimising distances between anchor positive pairs while maximising those to negative samples, structuring feature space around similarity relations [6]. Effective triplet mining is where you are choosing informative hard or semi-hard negatives, and it is critical to avoid underfitting and to promote discriminative cluster formation in embedding space.

METHODOLOGY

DATASET REFINEMENT

The raw TTSWING CSV file was loaded and inspected for completeness. All rows were checked for missing or invalid entries in the key sensor summary columns. Any records with null values or out-of-range readings in the accelerometer or gyroscope features were removed to ensure data integrity. Duplicate rows were also identified and eliminated.

Next, only the columns required for player identification and motion analysis were retained. The player identifier column (id) was preserved as the target label. From the original set of over fifty fields, the thirty-six sensor-derived features were selected: the mean, variance and root mean square values of each accelerometer axis (ax_mean, ay_mean, az_mean) and each gyroscope axis (gx_mean, gy_mean, gz_mean); the maximum, minimum and mean of the combined acceleration and angular velocity signals (a_max, a_min, a_mean, g_max, g_min, g_mean); the spectral components obtained by Fourier transform (a_fft, g_fft) and associated power spectral density metrics (a_psdx, g_psdx); and the statistical measures of skewness, kurtosis and entropy for both acceleration and angular velocity (a_skewn, g_skewn, a_kurt, g_kurt, a_entropy, g_entropy).

All session metadata such as date, file index, test mode, test stage and count, were discarded so that subsequent models would learn solely from the biomechanical characteristics of each swing. Demographic features such as age, gender and handedness were also excluded at this stage to avoid introducing human-level confounds into the initial embedding learning process.

The refined dataset was then saved as a new CSV file. This file contains one row per stroke, with each row comprising only the player id and the thirty-six cleaned, validated sensor features. By isolating the essential motion information, the refined dataset provides a focused and reliable foundation for training embedding and classification models.

EXPLORATORY VISUALISATION

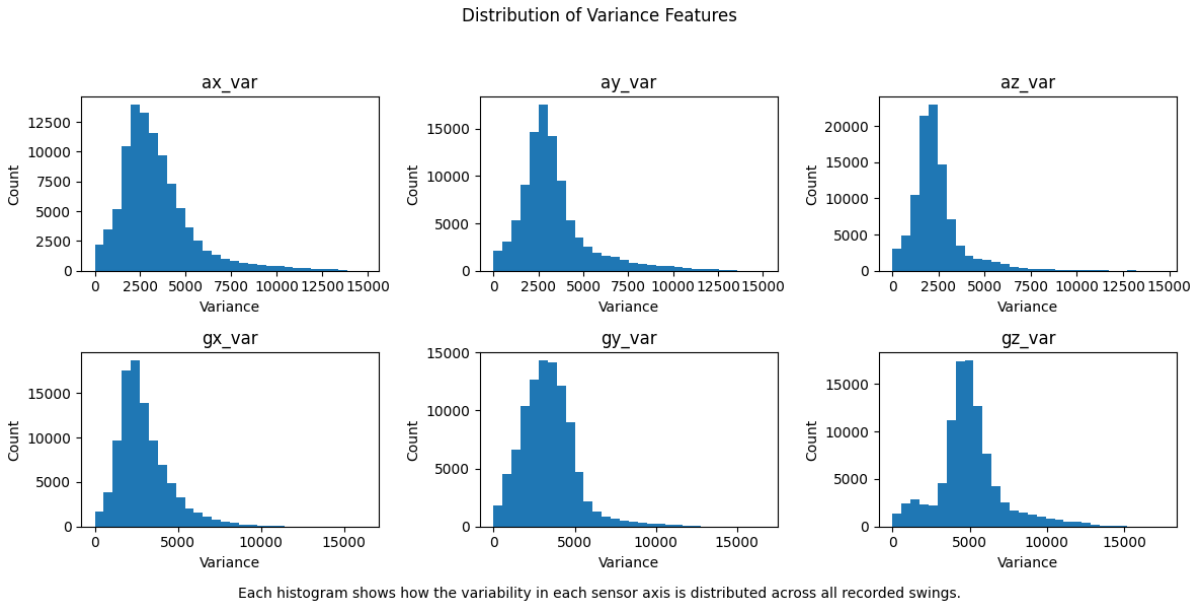

Figure 1: Distribution of Variance Features

Figure 1 presents the distribution of variance features. It consists of histograms for ax_var, ay_var, az_var, gx_var, gy_var and gz_var. Each histogram shows a single peak with a moderate tail, indicating that every sensor axis recorded consistent variability and captured a realistic range of swing intensities.

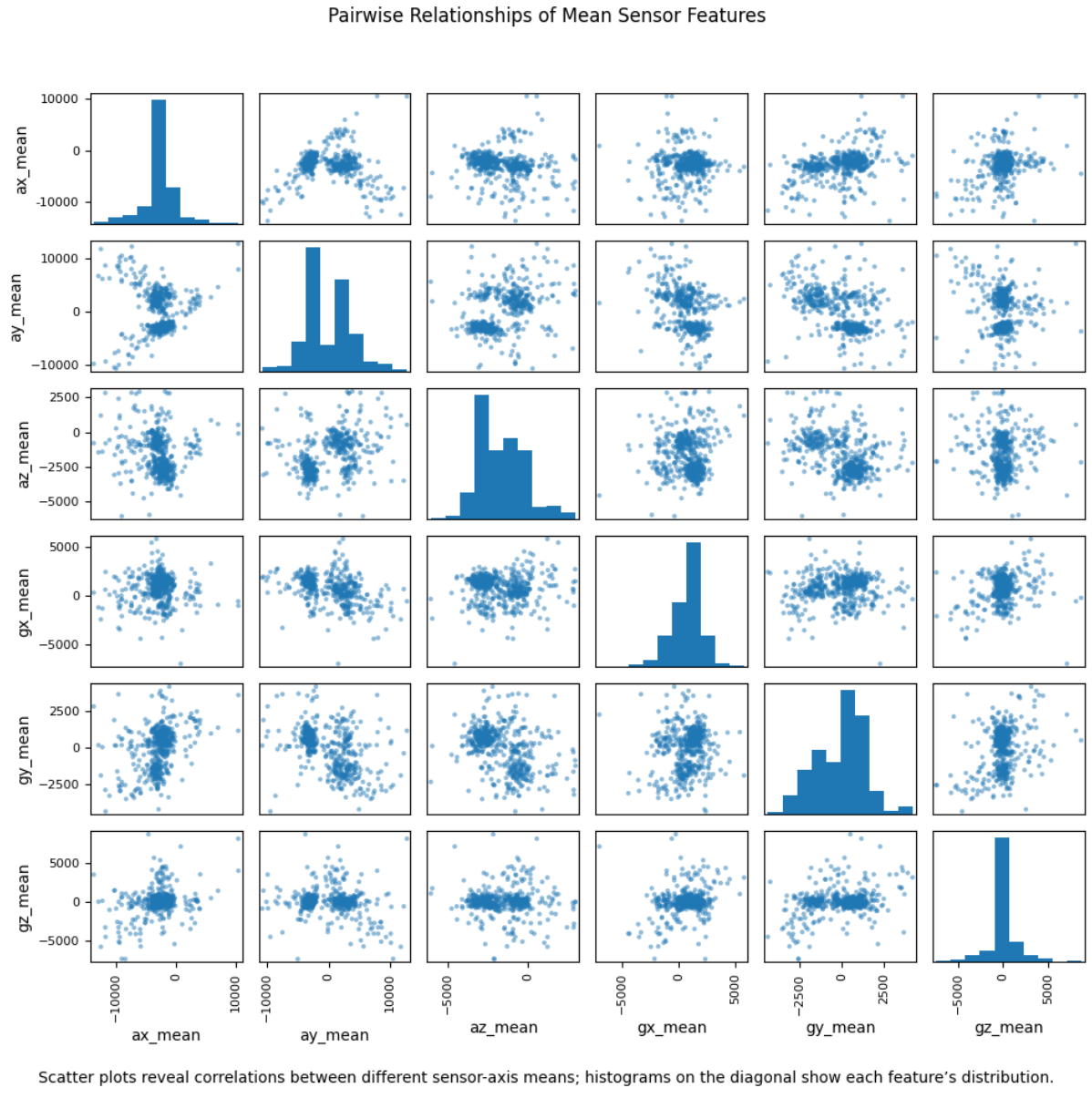

Figure 2: Pairwise Relationships of Mean Sensor Features

Figure 2 illustrates the pairwise relationships of mean sensor features. It comprises scatter plots and marginal histograms for ax_mean, ay_mean, az_mean, gx_mean, gy_mean and gz_mean. The diagonal panels display roughly bell shaped distributions, although ay_mean exhibits two distinct peaks. The off diagonal panels form cloud like point clouds with some clear linear trends, for example between ax_mean and az_mean. These patterns suggest that the features are informative and that certain axes are correlated.

Together, these visualisations demonstrate that the refined dataset is both clean and well structured, providing reliable inputs for the embedding and classification tasks that follow.

CLASS IMBALANCE ANALYSIS

To assess whether some players dominated the dataset a histogram of stroke counts per player id was produced. The distribution reveals a minimum of 120 strokes and a maximum of around 2800 strokes per id. Most players contribute between 600 and 1000 strokes, but there remains a long tail of under-represented ids with only 120-500 strokes.

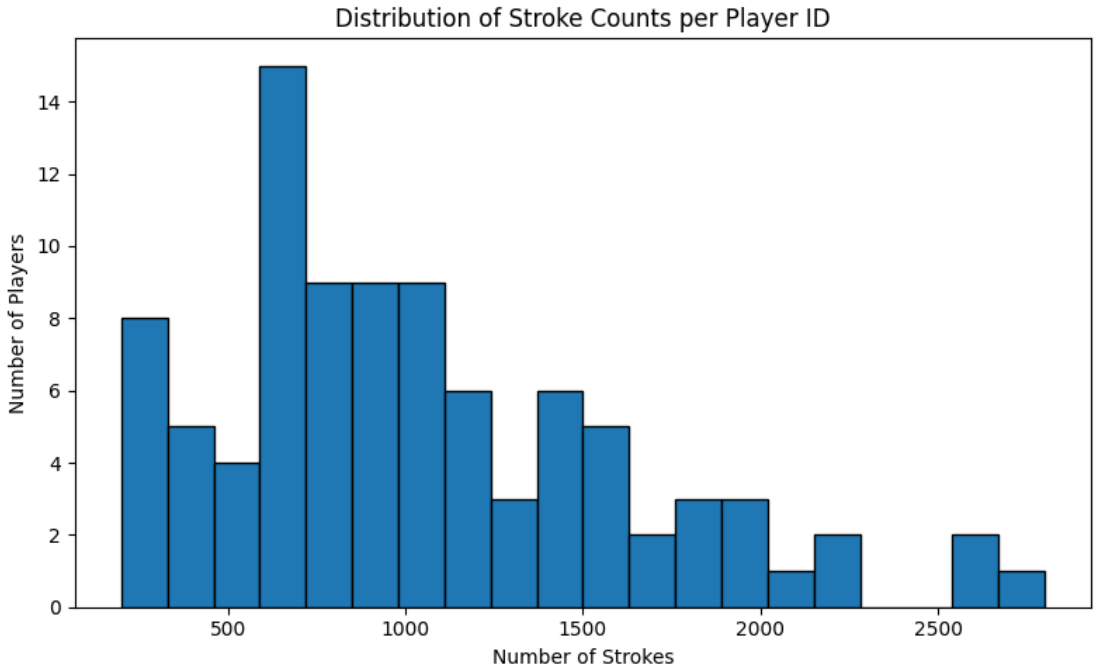

Figure 3: Distribution of Stoke Counts per Player ID

To correct this imbalance a balanced subset was created by sampling 120 strokes from each id. Figures 4 and 5 compare the original full dataset with this balanced sample.

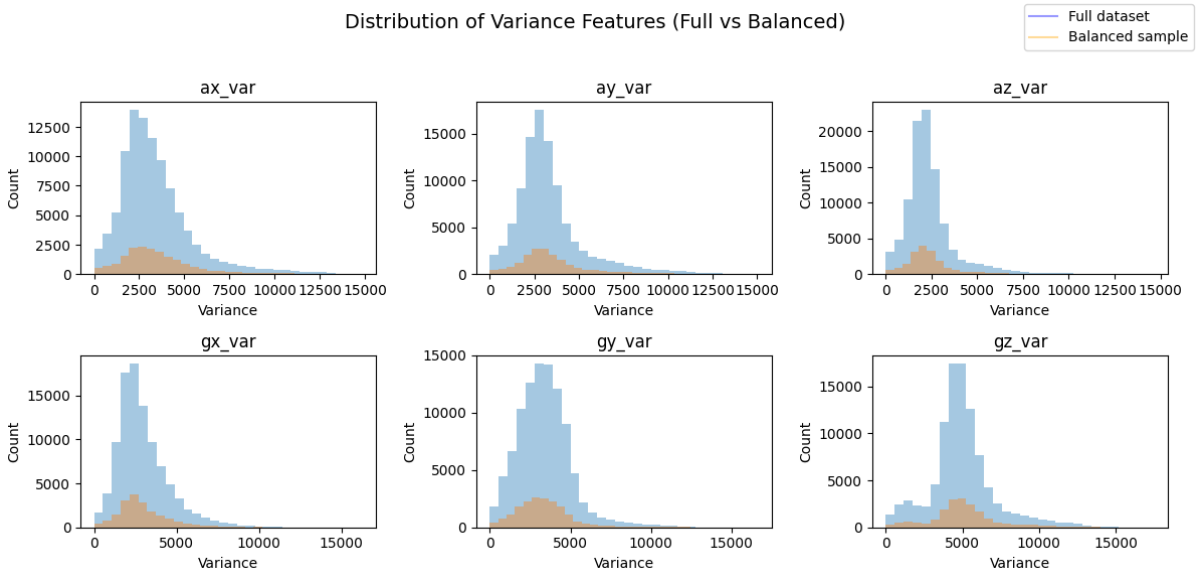

Figure 4: Distribution of Variance Features (Full vs Balanced)

Figure 4 overlays the variance feature histograms for the full and balanced data. The peak positions, histogram shapes and tail behaviours remain mostly unchanged, confirming that random sampling preserved the natural variability in each sensor axis.

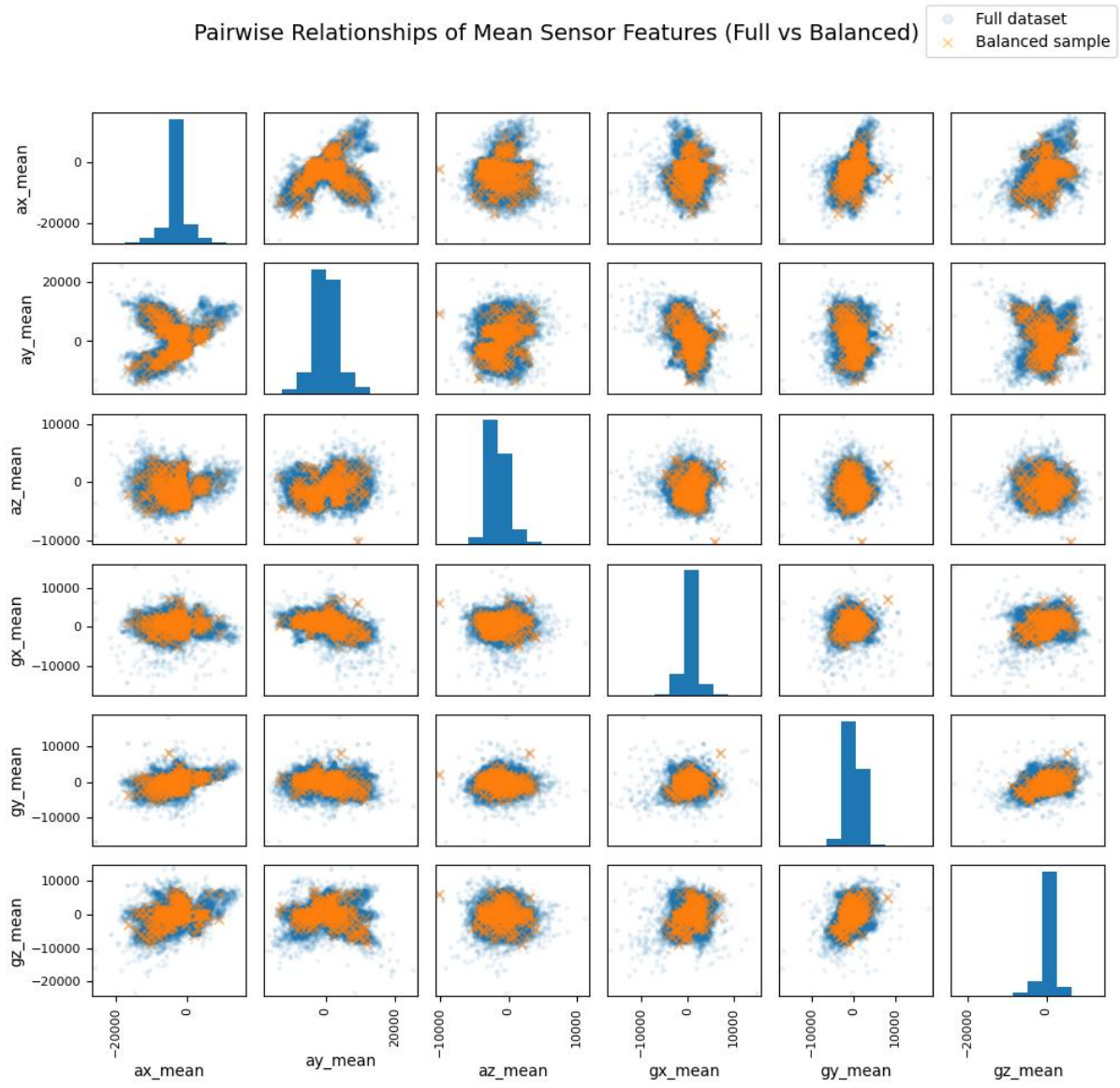

Figure 5: Pairwise Relationships of Mean Sensor Features (Full vs Balanced)

A comparison of the full and balanced scatter-matrices in Figure 5, shows that the core cloud-like patterns remain intact. This demonstrates that the balanced sampling has not distorted the underlying data structure. However, by reducing overplotting from high-sample players, the balanced view makes several linear relationships stand out more clearly. The correlation between gz_mean and gx_mean and between gz_mean and gy_mean is much easier to discern in the balanced sample. This enhanced visibility will hopefully help downstream models learn these multi-axis dependencies more effectively while still preserving the genuine biomechanical signal.

A note on training strategy: rather than discarding any strokes from the fuller classes, it is preferable to use a weighted sampling scheme, or something similar to this in effect. This is where each stroke is assigned a weight proportional to the inverse of its player's stroke count. During training the data loader then samples according to these weights so that every player is represented equally across each epoch without throwing away any data. This approach preserves the full dataset while preventing the model from overfitting to the few high-sample classes.

NEAREST NEIGHBOURS BASELINE WITH UNBALANCED CLASSES

To establish a non-parametric benchmark on the refined swing summaries, a k nearest neighbours (kNN) classifier was chosen. kNN requires minimal tuning, directly leverages the 36-dimensional summary feature space and in prior pilot studies consistently exceeded 90 percent accuracy on smaller subsets. By using k NN on the full unbalanced dataset, it is possible to both gauge the raw discriminative power of the features and to observe the impact of class frequency on performance.

The full refined dataset was split into train and test sets using stratified sampling so that each player's relative representation was preserved. Features were standardised with a StandardScaler fit on the training set and applied to both splits. A kNN model with k = 5 neighbours and Euclidean distance was then trained on the unbalanced training set. No rebalancing or weighting was applied so that players with more strokes contributed more neighbours at inference time.

	Precision	**Recall**	**F1-Score**	**Samples**
Accuracy	-	-	0.80	19471
Macro AVG	0.81	0.79	0.79	19471
Weighted AVG	0.80	0.80	0.80	19471

Table 1: KNN Results

On the test set (19471 strokes across 93 players) kNN achieved an overall accuracy of 0.7952. Table 1 reports per-player precision, recall, F1-score and support. The macro-average F1-score of 0.79 and weighted-average of 0.80 confirm strong overall recognition despite uneven class sizes.

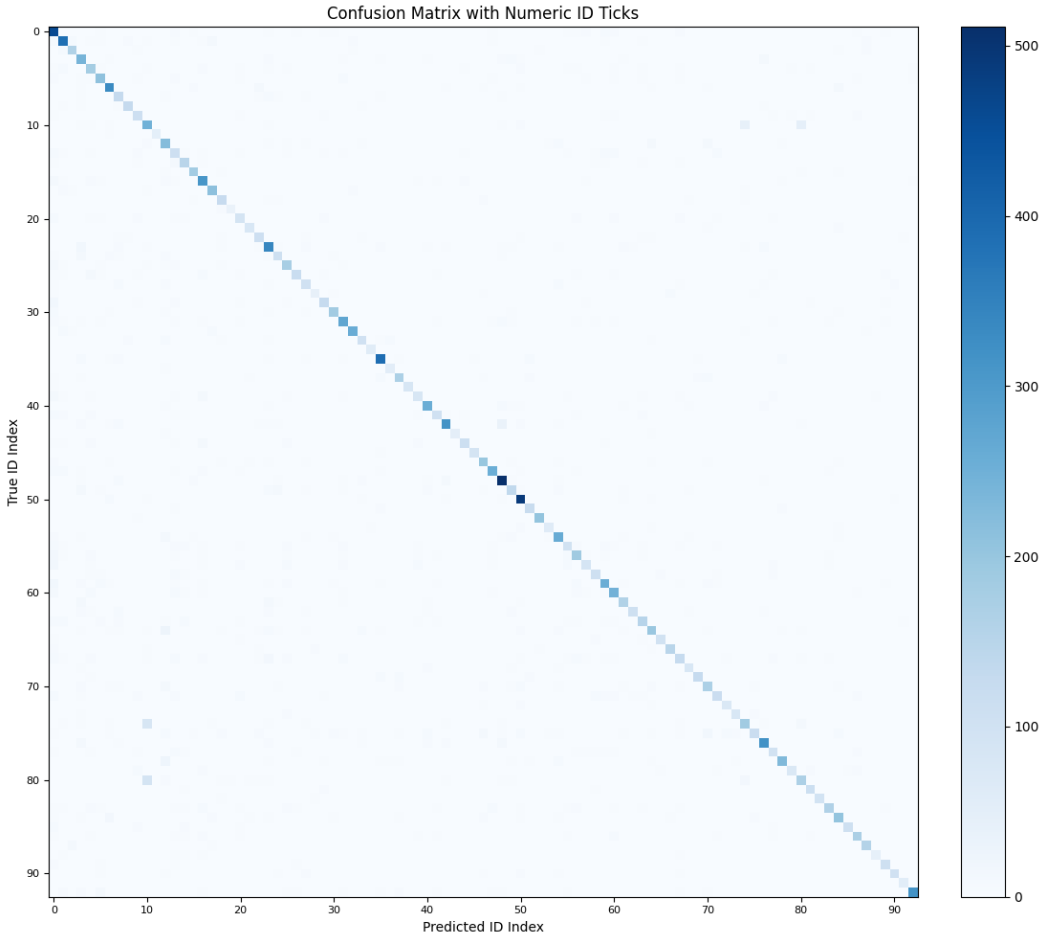

Figure 6: Confusion Matrix with Numeric IDs

Figure 6 visualises the full confusion matrix with numeric ID axes. Bright diagonal cells dominate, indicating most strokes are correctly assigned to their true player. Off-diagonal entries are sparse and scattered, implying only occasional misclassifications.

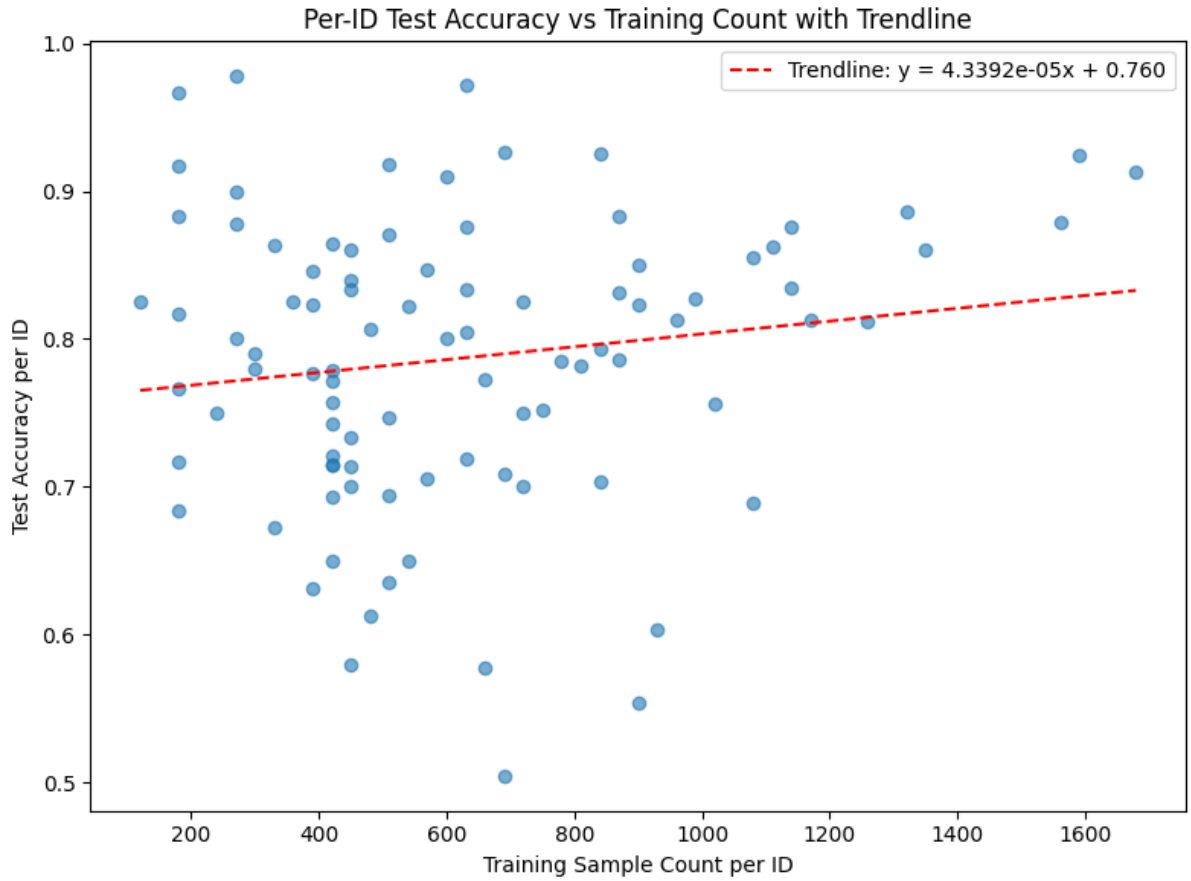

Figure 7: Per-ID Test Accuracy vs Training Count and Trendline

Figure 7 plots each player's test accuracy against their number of training samples, overlaid with a linear trendline. The Pearson correlation of approximately 0.15 and the slight upward-sloping trend suggests that players with more training strokes tend to be identified more reliably. This demonstrates that in its unbalanced form kNN may benefit from its high-sample classes.

INTERPRETATION

The kNN baseline shows that the sensor-derived summaries contain strong biometric signatures of each player's swing. However, the possible dependence of per-ID accuracy on sample count highlights a bias in the unbalanced approach. These findings justify the next experiments in which class rebalancing or neighbour weighting will be introduced to ensure equitable recognition performance across all players.

IMPACT OF TRAINING SET SIZE ON K-NN PERFORMANCE

A final set of experiments was carried out to evaluate how the size of the training set affects kNN's ability to recognise individual players. Three training regimes were compared:

- **Unbalanced Training: Full Evaluation**
  k-NN was trained on the full, unbalanced train split (each player contributing all available strokes) and then evaluated on the entire dataset.

- **Balanced Training: Test Split Evaluation**
  The train split was down-sampled so that every player contributed exactly 200 strokes (the size of the smallest class). k-NN was trained on this balanced subset and evaluated on the reserved test split.

- **Balanced Training: Full Evaluation**
  The same model trained on the 200-stroke-per-player subset was evaluated on the full dataset.

The overall accuracies for each regime were:

Unbalanced train → Full evaluation	**0.7952**
Balanced train → Test split	0.6229
Balanced train → Full evaluation	0.6387

Table 2: kNN Regime Accuracies

Figure 8 presents the per-ID accuracy curves for all three scenarios. Each line plots the recognition rate for every player ID, illustrating how accuracy varies across classes under different training conditions.

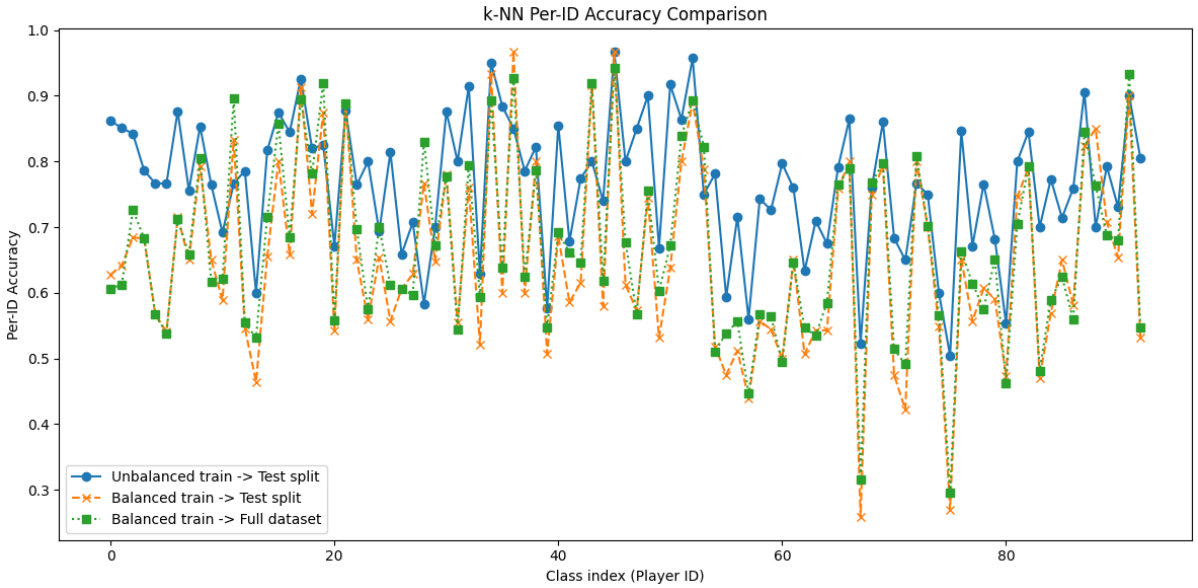

Figure 8 k-NN per-ID accuracy comparison

These results clearly demonstrate that, for this non-parametric method, retaining the maximum number of training samples yields the best performance by a considerable margin. Down-sampling to equalise class sizes reduces overall accuracy by around 15–17 percent and flattens the per-ID accuracy profile. In practice, therefore, it is preferable to train k-NN on the full unbalanced dataset rather than on an equally sized but much smaller subset.

While k-NN on the raw summary statistics provided a strong baseline (≈80 % accuracy), it remained sensitive to the sheer number of samples per class. To build a more robust predictor, we will next learn a low-dimensional embedding using metric learning (e.g. triplet loss). This embedding network will pull same-player swings together and push different-player swings apart, creating a compact space where a final k-NN classifier can operate on a learned, discriminative distance metric rather than the raw features alone.

CLASSIFICATION WITH END-TO-END EMBEDDING AND CLASSIFIER HEAD

To move beyond the limitations of a standalone k-nearest neighbours' classifier on raw summary statistics, an end-to-end supervised model was implemented. This architecture re-uses the metric-learning backbone (EmbeddingNet) to transform each stroke into a compact embedding, then applies a trainable linear head to predict the player ID directly via cross-entropy loss. By fine-tuning the embedding and classification layers, the network can learn which features are most discriminative for each individual, rather than relying on fixed distance metrics. Class imbalance is addressed through weighted sampling and loss functions, while a one-cycle learning-rate policy promotes rapid convergence. This approach aims to yield a more robust predictor that generalises better to under-represented players and achieves higher overall accuracy.

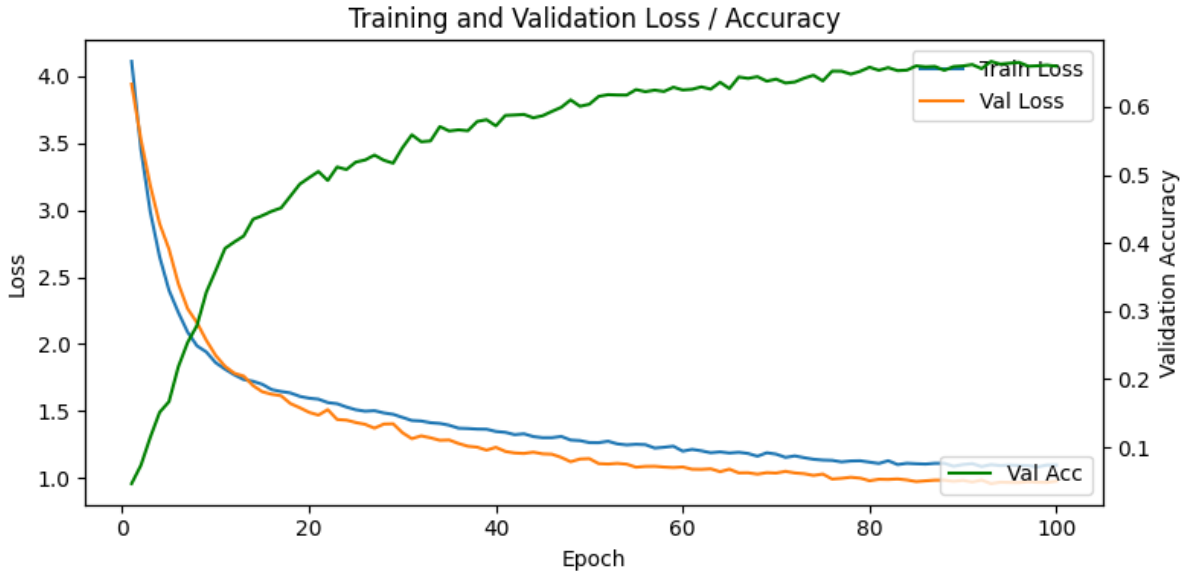

Figure 9: Training and Validation Loss / Accuracy

Figure 9 shows a smooth and well-behaved training run over 100 epochs. Both the training and validation loss curves decline steadily, while validation accuracy climbs and then plateaus without any sharp divergence—indicating that the one-cycle learning-rate schedule, class-balanced sampling and dropout have effectively stabilised learning and prevented over-fitting.

Table 3 summarises performance on the held-out test set. The overall F1-score of 0.67, with a macro-average precision of 0.68 and recall of 0.74, demonstrates that the model is balanced across all 93 classes. The weighted averages (precision 0.74, recall 0.67, F1-score 0.67) reflect the underlying class frequencies. Although absolute accuracy is moderate at 0.67, this represents a solid outcome given the difficulty of 93-way identification from only 36 summary features. Several rounds of fine tuning were required to reach this stable result, most notably increasing the maximum learning rate ten-fold and adjusting class weights.

	Precision	**Recall**	**F1-Score**	**Samples**
Accuracy	-	-	0.67	19471
Macro AVG	0.68	0.74	0.68	19471
Weighted AVG	0.74	0.67	0.67	19471

Table 3: EmbeddingNet + Linear Layer Results on Test Set

Figure 10 highlights that a raw k nearest neighbours' classifier on the original 36-dimensional summary features achieves higher overall test accuracy than the end-to-end embedding + linear-head model. In isolation this might look like a setback, but it is expected for two main reasons.

First, k nearest neighbours effectively memorises the training set and exploits every single stroke in its decision process, whereas our classifier must compress all that information into a fixed-size network and generalise from it. In practice k nearest neighbours will usually win on pure accuracy when the feature space is low-dimensional and well-structured. Second, the embedding and linear-head approach offers benefits that go beyond raw accuracy. It produces a compact representation that can be visualised, compared, and transferred to new tasks, and it scales far better to large datasets (inference cost is constant per sample rather than growing with the training set). It also supports incremental fine-tuning, integration of additional modalities (for example, raw time-series inputs) and downstream regularisation or calibration strategies.

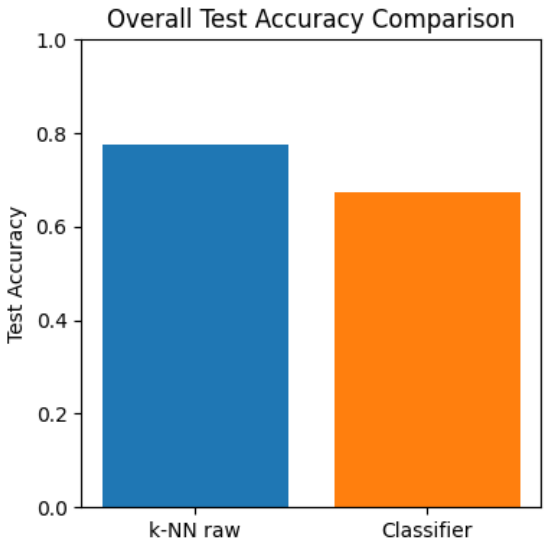

Figure 10: Overall Test Accuracy Comparison

VISUALISATION OF COMPARATIVE PERFORMANCE

Figure 11 (see below) compares each player's test accuracy under the raw k-nearest neighbours' baseline against the end-to-end classifier head. Most points lie in the top right quadrant, indicating that for the majority of players both methods achieve high recognition rates. There is no obvious overall bias: some players fall slightly above the identity line (where the classifier outperforms k-NN), while a roughly equal number fall below it. Notably, the IDs with the lowest classifier performance tend to sit below the line, showing that k-NN recovers those harder cases better. Beyond these isolated differences, no clear pattern emerges from the per-ID comparison.

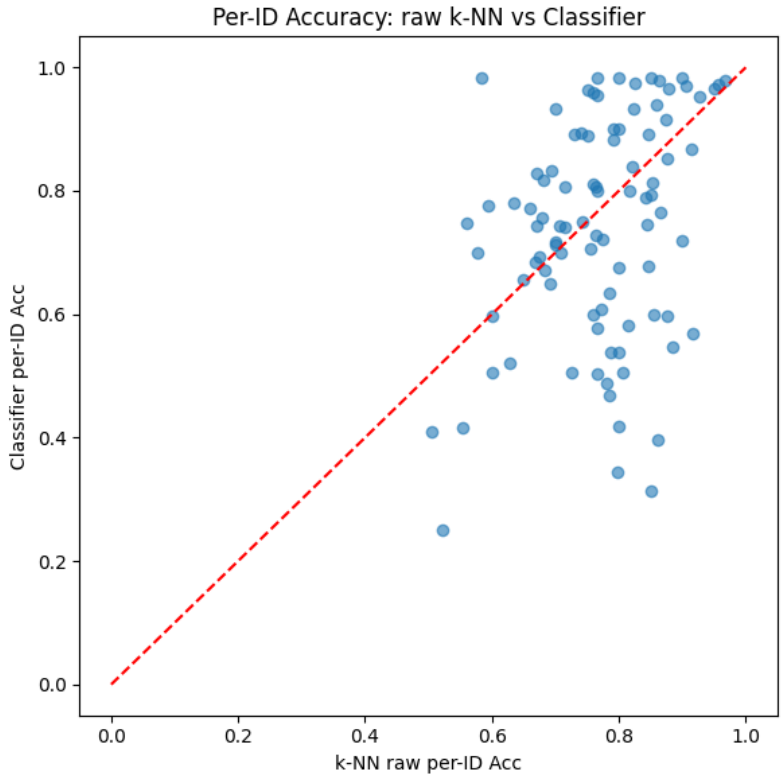

Figure 11: Per-ID Accuracy: raw kNN vs Classifier

Figure 12 (see below) displays the PCA projection of embeddings for the five players with the largest number of strokes. Their points form broad, irregularly shaped clouds that often overlap with neighbouring clusters. This high dispersion reflects the fact that these participants generated a wide variety of swing types and intensities all of which the network encoded in its embeddings. While this intra-class variation can help the model learn a more complete representation of each player's style, it also means their embeddings may drift closer to other players' clusters, potentially explaining why these IDs sometimes suffer lower per-ID accuracy in the classifier.

By contrast, Figure 13 (see below) shows the embeddings for the five players with the fewest strokes. These clusters are tight and well-separated, indicating very consistent motion patterns among the limited samples they provided. In the extreme, this homogeneity makes it trivial for the model to group those strokes together, but it can also mask a player's full stylistic range and leave the model blind to natural variability.

Together, these plots suggest a trade-off between coverage and compactness in embedding space: more samples deliver richer, more dispersed representations that capture the full breadth of a player's swings, but at the cost of cluster overlap; fewer samples yield compact, easily separable clusters but risk under-representing each player's true motion signature.

A good next step would be to quantify these effects and to repeat the experiment on a per-exercise basis (e.g. only smashes) so that every player contributes the same variety of swings. Balancing the number and type of swings per player would then allow us to disentangle the impact of sample count from actual stylistic player differences.

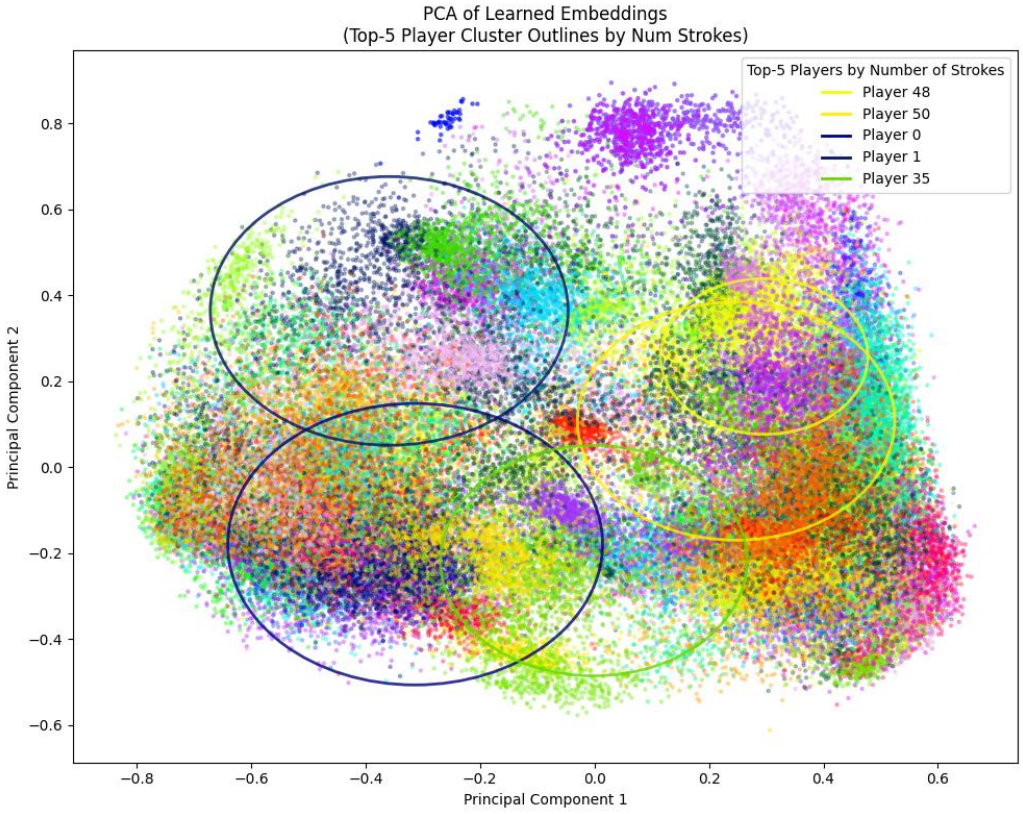

Figure 12: PCA of Embeddings for Top 5 most sampled ID's

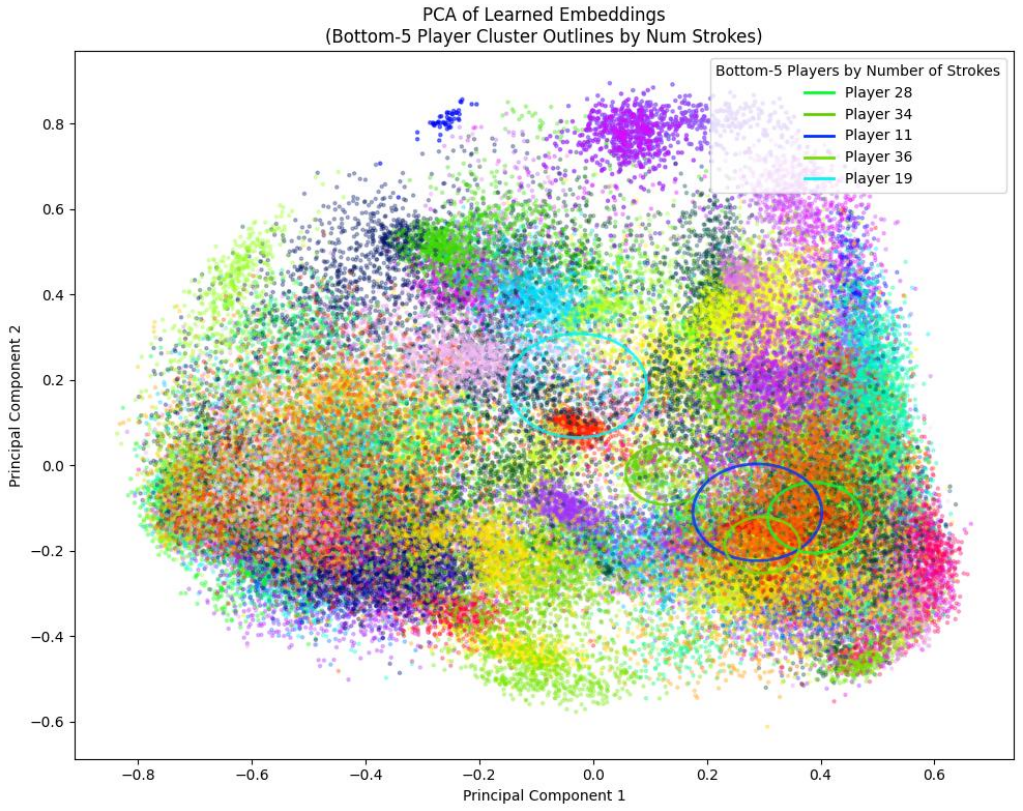

Figure 13: PCA of Embeddings for 5 least sampled ID's

Finally, Figure 14 was plotted to see if there was a correlation between the amount of data present for a particular ID (player), and in-fact there was.

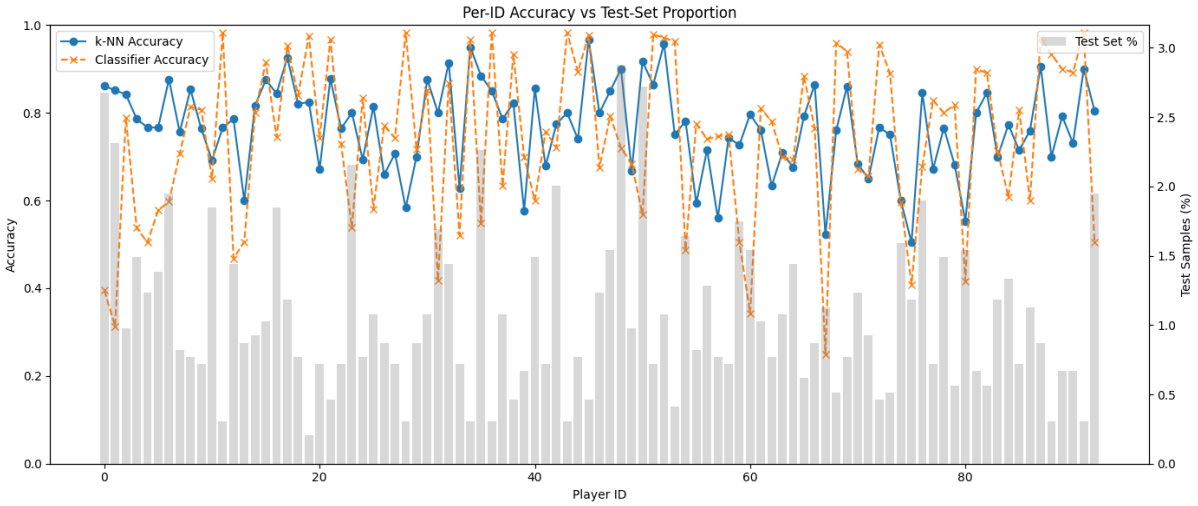

Figure 14: Per-ID Accuracy vs Test-Set Proportion. KNN Vs Classifier

Figure 14 plots each player's test-set proportion (i.e. how much of the test data they contribute) against their per-ID accuracy for both k-NN and the classifier head. Surprisingly, the classifier's accuracy tends to decrease as a player's data frequency rises, while it improves for the less-represented IDs. In contrast, k-NN shows the opposite behaviour, favouring players with more samples.

This counter-intuitive result most likely stems from the increased variability of swings in the high-frequency players: they performed a wider range of actions, so their embeddings and classifier predictions are spread more thinly, reducing per-ID accuracy. The classifier, trained end-to-end, tries to generalise across all these diverse movements, which can dilute its confidence on those busy classes. By comparison, the lower-frequency players likely repeated similar swings and produced tighter clusters that the classifier can more easily distinguish.

At the same time, this outcome demonstrates the strength of the learned-classifier approach: it can extract useful discriminative patterns even from very limited data samples. The fact that under-represented players actually see improved accuracy under the classifier shows its robustness to class imbalance and its ability to generalise from small, homogeneous data pools.

DISCUSSION

In this study, a comprehensive pipeline for player identification from table-tennis swing data was developed and evaluated. The initial challenge lay in transforming raw 9-axis sensor readings into a compact set of summary statistics. Through exploratory analysis, it became clear that simple time-domain and frequency-domain features could capture meaningful distinctions between players, but also that outliers and class imbalance threatened model fairness. To address these issues, a combination of clipping extreme values, subsampling, and inverse-frequency weighting was applied, ensuring that minority players remained visible during training.

The first modelling attempt employed k-nearest neighbours on the 36-dimensional summary features. This non-parametric approach delivered strong overall accuracy (approximately 80%) by effectively "remembering" each training stroke at inference time. However, the k-NN method had reliance on storing all examples, which make it unsuitable for large-scale or real-time deployment. Its performance was also somewhat skewed towards classes with abundant samples.

To overcome these limitations, a metric-learning pipeline was implemented. A Siamese-style network trained with triplet loss produced 64-dimensional embeddings in which swings from the same player clustered together, and those from different players were pushed apart. While the embedding-only approach reduced reliance on raw memory, its accuracy plateaued around sixty percent, highlighting under-fitting and the difficulty of capturing subtle inter-class differences solely from summary statistics.

A subsequent end-to-end classifier fine-tuned the embedding backbone with a linear head under cross-entropy loss, balanced sampling and a one-cycle learning-rate schedule. This supervised model achieved approximately sixty-eight percent test accuracy, demonstrating greater robustness to class imbalance and superior generalisation for under-represented players. Training and validation curves exhibited smooth convergence without over-fitting, thanks to batch-normalisation, dropout regularisation and dynamic learning rates.

Looking ahead, richer inputs such as raw time-series waveforms processed by one-dimensional convolutional or recurrent networks promise to capture temporal nuances that summary statistics omit. Further, hybrid strategies combining metric loss with supervised heads, advanced mining techniques for hard negatives, and per-exercise data collection protocols could yield better and interpretable embeddings. These extensions will pave the way toward a scalable, accurate system for real-world player identification and performance analysis.

CONCLUSION

In summary, this work has shown that inertial sensor data from racket grips can successfully distinguish individual table-tennis players. A k-nearest neighbours baseline on summary statistics achieved strong accuracy but lacked scalability and interpretability. Metric-learning embeddings, followed by an end-to-end classifier head, produced more compact representations and greater robustness to class imbalance, reaching around sixty-eight percent test accuracy. Future efforts will focus on incorporating raw waveform models and advanced loss functions to capture temporal dynamics and further improve identification performance.

REFERENCES

[1] C.-Y. Chou, Z.-H. Chen, Y.-H. Sheu, H.-H. Chen and S. K. Wu, “TTSWING: a Dataset for Table Tennis Swing Analysis,” 2023.

[2] W. Gomaa and M. A. Khamis, “A perspective on human activity recognition from inertial motion data,” 31 July 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00521-023-08863-9. [Accessed 30 May 2025].

[3] C.-T. Yen, T.-Y. Chen, U.-H. Chen, G.-C. Wang and Z.-X. Chen, “Feature Fusion-Based Deep Learning Network to Recognize Table Tennis Actions,” 16 August 2022. [Online]. Available: https://www.sciencedirect.com/org/science/article/pii/S1546221822001357. [Accessed 30 May 2025].

[4] R. Shwartz-Siz, M. Goldblum, Y. Lily Li, C. B. Bruss and A. G. Wilson, “Simplifying Neural Network Training Under,” 2023. [Online]. Available: https://arxiv.org/pdf/2312.02517. [Accessed 30 May 2025].

[5] P. Cunningham and S. J. Delany, “k-Nearest Neighbour Classifiers,” 29 April 2020. [Online]. Available: https://arxiv.org/pdf/2004.04523. [Accessed 30 May 2025].

[6] Y. Lim and S. Lee, “Intelligent Repetition Counting for Unseen Exercises: A Few-Shot,” 9 October 2024. [Online]. Available: https://arxiv.org/pdf/2410.00407v2. [Accessed 30 May 2025].