Astrometric Binary Classification Via Artificial Neural Networks

Joe Smith [ID]

Boyce Research Initiatives and Education Foundation, San Diego, CA 92106, USA; jsmithsd.ohs@gmail.com

ABSTRACT

With nearly two billion stars observed and their corresponding astrometric parameters evaluated in the recent *Gaia* mission, the number of astrometric binary candidates have risen significantly. Due to the surplus of astrometric data, the current computational methods employed to inspect these astrometric binary candidates are both computationally expensive and cannot be executed in a reasonable time frame. In light of this, a machine learning (ML) technique to automatically classify whether a set of stars belong to an astrometric binary pair via an artificial neural network (ANN) is proposed. Using data from *Gaia* DR3, the ANN was trained and tested on 1.5 million highly probable true and visual binaries, considering the proper motions, parallaxes, and angular and physical separations as features. The ANN achieves high classification scores, with an accuracy of 99.3%, a precision rate of 0.988, a recall rate of 0.991, and an AUC of 0.999, indicating that the utilized ML technique is a highly effective method for classifying astrometric binaries. Thus, the proposed ANN is a promising alternative to the existing methods for the classification of astrometric binaries.

1. INTRODUCTION

A binary star system is where two or more stars are gravitationally bound, orbiting about a common center of mass. Astrometric binaries are binary star systems inferred by considering the astrometric parameters (positions, proper motions, parallaxes, and radial velocities) of each component star. Astrometric binaries provide significant value, as other than being used to map the cosmos at large, their astrometric parameters can be used to determine the masses of the component stars (Zavada & Píška (2020)), examine open cluster and star field kinematics (Gruner et al. (2023)), identify limits on the dark matter density present in massive compact halo objects (Yoo et al. (2004)), and can be utilized in general gravitational probes (Hernandez et al. (2024)). Hence, it is very important that binary star catalogs are constructed in order to be applied to the research of a variety of other astrophysical phenomena.

ML techniques, and deep neural networks (DNNs) in particular, have been applied to astronomy at large in recent years, prompting many new discoveries and questions (e.g., Lin et al. (2020b); Li et al. (2020); Szklenár et al. (2022)). One especially popular method of DNNs, ANNs, have been commonly applied to problems in-volving signal analysis and random function approximations. ANNs have been extensively applied to classification problems pertaining to astronomy, such as the morphological classification of galaxies (e.g., Lahav et al. (1996); Banerji et al. (2010)), the classification of galaxy spectra (Storrie-Lombardi et al. (1992); Folkes et al. (1996)), the classification of stellar spectra (von Hippel et al. (1994); Bailer-Jones et al. (1998)), and the selection of pulsar candidates from radio surveys (Eatough et al. (2010)). However, while astronomers continue to automate the classifications and predictions of other astrophysical phenomena, they have largely overlooked applying such ML techniques to astrometry.

In recent times, sky surveys and space missions have obtained vast amounts of data on individual stars and their relative positions and motions. Namely, the *Gaia* mission (Gaia Collaboration et al. (2016)) launched by the European Space Agency (ESA) has observed over 1.5 billion stars and determined their respective astrometric parameters with unprecedented precision. Due to the incredible scale of this mission, a surplus of new binary star candidates have been introduced for classification. In light of this, astronomers have employed computational techniques to compute binary catalogs by considering the astrometric parameters of potential bi-

nary star pairs (e.g., El-Badry et al. (2021); Medan & Lépine (2023)). However, not only are these methods restricted to only limited subsets of *Gaia* data, they are also inefficient at classifying a group of stars as binary, often requiring many intermediate steps before successfully classifying an input set of stars. Consequently, it has proven to be very difficult to analyze the entire *Gaia* database to construct an appropriate binary catalog. Therefore, an automated technique for the classification of astrometric binaries is required.

In this paper, an ANN model is proposed for the automated classification of astrometric binaries. The proposed ANN aims to solve the problem posed by the present methods of the classification of astrometric binaries in order to maximize the potential of *Gaia* data releases and future surveys. This work also offers an opportunity to compare human classifications to those from automated ML algorithms on a large scale. If proven to be as effective as human classifications of astrometric binaries, the proposed ANN could save significant computing time and costs for future surveys and studies.

The rest of this paper is organized as follows. In Section 2, the data and the methods for data preprocessing are presented, while Section 3 describes the proposed ANN. The results and associated discussions are given in Section 4. The paper closes with Section 5, in which the conclusions about the efficacy of the proposed ANN are given.

2. DATA AND METHODS

Given the scale and precision of the *Gaia* mission, this work will utilize data from *Gaia* Data Release 3 (DR3; Gaia Collaboration et al. (2023)) to train and test the proposed ANN.

2.1. *Feature Selection*

In the construction of the data set, the radial velocities (RVs) of each binary are neglected. This is because of the more than 1.5 billion stars observed by the *Gaia* mission, only 33 million have their RVs measured. As a result, the proposed ANN will only be trained on a subset of the overall astrometric parameters of each binary: the positions, proper motions (μ_1, μ_2 mas yr^{-1}), and parallaxes (ϖ_1, ϖ_2 mas). Here, μ_1 and μ_2, and ϖ_1, ϖ_2, indicate the proper motion and parallax values of the primary and secondary component stars, respectively. The declination (δ_1, δ_2) and the right ascension (α_1, α_2) of the component stars of each binary cannot be meaningful alone, but they can be used to compute their associated angular ($\theta°$) and physical (s au) sepa-

rations. The angular separation between any two component stars is defined as

$$\frac{\Theta}{''} = 2\sin^{-1}\left(\sqrt{\sin^2\frac{\Delta\delta}{2} + \cos\delta_1\cos\delta_2\sin^2\frac{\Delta\alpha}{2}}\right), \quad (1)$$

where $\Delta\delta = \delta_1 - \delta_2$ and $\Delta\alpha = \alpha_1 - \alpha_2$. The angular separation is converted into degrees by $\theta = \frac{\Theta}{3600}$. The physical separation is computed as

$$\frac{s}{\text{au}} = 1000\frac{\Theta}{\varpi}, \quad (2)$$

serving as an accurate approximation despite not being equivalent to the complete 3D separation (El-Badry & Rix (2018)).

These six features (inputs) will be considered to train and test the ANN. The aim is to effectively train the ANN to map the features of binary star candidates to their correct class: true binaries (positives) or visual binaries (negatives).

2.2. *Numerical Data*

Many binary star catalogs have been computed utilizing the data available from *Gaia* data releases. One of the most extensive is provided in El-Badry et al. (2021), where 1.1 million highly probable binary star candidates are calculated within 1 kpc of the Sun from *Gaia* DR3. In deriving the catalog, conditions are imposed on certain astrometric parameters of each candidate component star: the parallax, proper motion, and physical separation values. Most notably, the parallax and proper motion values of potential binary star candidates are calculated such that they are within an appropriate relative range in comparison to their potential companion star. In other words, the catalog is derived on the basis of similar proper motion and parallax values. In regard to physical separations, the component stars must have a physical separation of less than 1 pc, thereby accounting for a range of binary star systems, such as wide binaries. This threshold is chosen since a very limited binary population is expected to exist at physical separations larger than this.

The catalog derived by El-Badry et al. is sifted for both highly probable true binaries and visual binaries, which are commonly referred to as "chance-alignments." This is done by sifting the computed chance-alignment rates, R, between $0.9 < R < 1.1$, where $R \sim 1$ is a highly probable visual binary, and $R < 0.1$, where $R \sim 0$ is a highly probable true binary. This yielded 1.5 million highly probable true and visual binaries, where true binaries were then assigned a "binary class" value
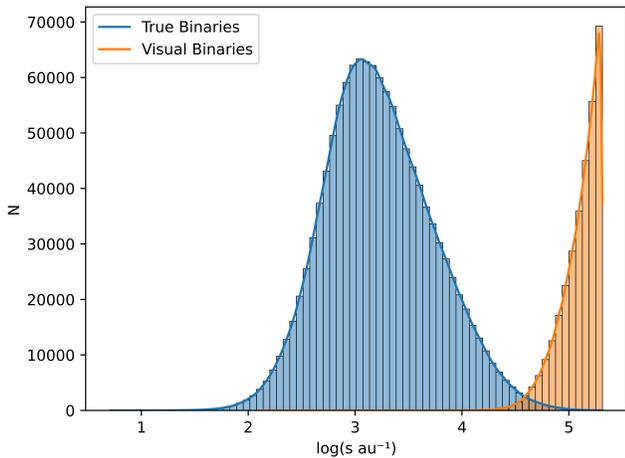
Figure 1. Histogram of true and visual binaries in the sorted catalogue as a function of their physical separation. The range of values fall between $0.70 < \log(s\,\mathrm{au}^{-1}) < 5.31$, obeying the initial condition of $s < 206265$ au.
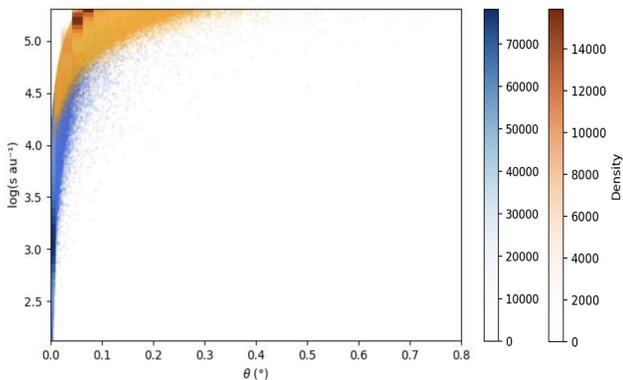
Figure 2. Physical separations plotted against their corresponding angular separations. Overplotted for true (blue) and visual (orange) binaries are 2D density histograms.

of "0" and visual binaries a binary class value of "1". The parallaxes, proper motions, and physical and angular separations of each true and visual binary are then extracted into a separate data set. Since the selected features do not include the relative distances of each binary candidate from the Earth or the Sun, the selection criteria initially imposed will not restrict the applicability of the trained ANN to arbitrary relative distances.

2.3. *Class Imbalance*

Considering the range of potential values for each astrometric parameter associated with either a true or visual binary, it is expected that there will be some degree of class imbalance present in the data. Classification algorithms often suffer from imbalanced data due

to the skew that is present towards one class. As a result, classification algorithms are easily affected by an imbalanced data set, which commonly leads to a poor classifier despite an increase in algorithm performance (Longadge & Dongre (2013); Johnson & Khoshgoftaar (2019)). The level of class imbalance is measured as the ratio between true binaries to visual binaries, which is called the imbalance ratio (IR). A popular standard defines a high class imbalance when the IR ranges from 100:1 to 10,000:1 (He & Garcia (2009)). There is an IR of 4:1 in the data set used in this work, indicating that the data set utilized will not significantly inhibit or inflate model classification scores considering the shape of the true and visual binary classes.

Figure 1 shows a histogram of each true and visual binary as a function of their physical separation. As can be seen in this figure, the physical separation distributions of each class indicate that the data set is strongly defined by the physical separation feature. Comparing the physical separation feature against the angular separation feature in Figure 2 yields a similar result, where there is similar clustering found in the equivalent physical separation regions. However, the binary classes are not as well defined by their angular separations as by their physical separations, where the true and visual binary classes have mean angular separations of $0.0023° \pm 0.0060°$ and $0.065° \pm 0.039°$, respectively. The implications for ML model performance as a result of the strong class definition given by the physical separation feature will be explored in greater depth in Section 4.4.

2.4. *Data Preprocessing*

The data set (1.5 million binaries) is split into three parts: training set (60%), validation set (20%), and test set (20%). The training set is used to fit the proposed ANN, while the validation set provides the information necessary to tune the selected hyperparameters of the ANN. Since there is limited class imbalance present in the data set, a data augmentation approach to minimize the effects of class imbalance during the training, validation, and testing phases is unnecessary.

3. ARTIFICIAL NEURAL NETWORK

There are many ML techniques available to apply to the problem of classifying binary star systems. For this work, an ANN is chosen due to the very large numerical data set that is used, which is ideal for training DNNs compared to, for example, support vector machines (SVMs; Hearst et al. (1998)). Compared to other DNNs, an ANN is also likely to be preferred for the classification of astrometric binaries. In particular, classification of via a convolutional neural network (CNN;
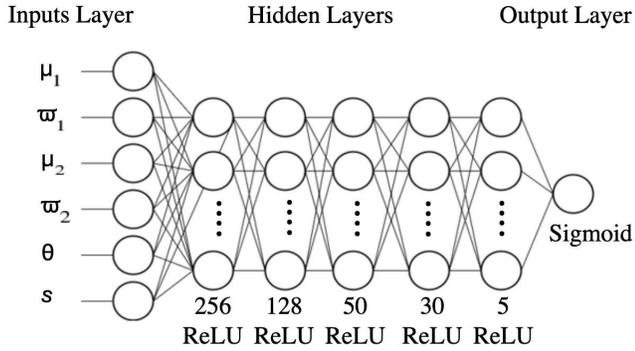
Figure 3. The architecture of the proposed ANN, where the inputs are the six features associated with the input binary candidate (Section 2.1). See Table 1 for the model architecture of the ANN in detail.

Table 1. Hyperparameters of the ANN model

Parameter	Tested Values	Selected Value
Architecture		
Number of dense layers	$[1 - 10]$	5
	ReLU	
Dense activation function	Sigmoid	ReLU
	Tanh	
Number of neurons per layer	$[1 - 1000]$	Fig. 3
Optimization		
Mini-batch size	$[16 - 2048]$	512
Learning rate	$[10^{-2} - 10^{-5}]$	2.5×10^{-4}
Optimizer		Adam

LeCun et al. (2015)) may not be feasible due to the limited information provided by images. However, additional work is necessary to confirm that CNNs cannot reliably classify astrometric binaries.

3.1. *Architecture of the Artificial Neural Network*

The proposed ANN was developed using the Keras API built within TenserFlow (Abadi et al. (2015)), an open source library for ML and deep learning. The architecture of this ANN can be seen in Figure 3.

The astrometric parameters in the sorted catalogue are essentially floating point numbers, which can be used as input parameters in the ANN without requiring any manipulation. Therefore, a simple dense layer with a sigmoid activation was applied after the hidden (dense) layers resolved a numerical input via rectified linear unit activation functions (ReLU; Nair & Hinton (2010)). This outputs a probability that determines whether or not the input binary star system is a visual or true binary under an operating threshold. The selected threshold for this ANN was the standard 0.5. If the output provided by the sigmoid activation function is greater than 0.5, then the input binary star system is classified as a visual binary, whereas if the output is less than 0.5, then the input binary star system is classified as a true binary. The classifications were compared with the ground truth of labeled classes using the binary cross entropy loss function (Hinton & Salakhutdinov (2006)) and batch stochastic gradient descent (SGD) with the Adam optimizer (Kingma & Ba (2014)).

Cross-validation (CV) methods are typically used to minimize the classification error of the DNN by providing information on the required tuning of the model hyperparameters (Hastie et al. (2009); Arlot & Celisse (2010)). Of the many CV methods available, this work employs a method known as hold-out CV due to the nature of the objectives of this work. The potential for overfitting due to hold-out CV is minimized by optimizing the hyperparameters of the learning algorithm (Dietterich (1995)).

3.2. *Model Optimization*

Hyperparameters of mini-batch size, number of epochs, and learning rate were optimized for more efficient model training. The model performed best when a relatively small mini-batch size of 512 and a Keras learning rate of 0.00025 were used. To prevent overfitting, a regularization step of early stopping was applied (Caruana et al. (2000)). Via a callback of validation loss, the early stopping function monitors the change of the validation loss value before stopping the training process and saving the best model weights. The number of epochs without improvement after which training will be stopped by the early stopping function is referred to as "patience." In this work, a patience of 20 and a minimum required change of 10^{-4} are utilized. Table 1 presents the tested and optimal hyperparameters of the ANN.

3.3. *Performance Measure*

In order to evaluate the performance of the DNN, several classification scores should be considered, such as the accuracy, the precision, the recall, and F1 score (Lin et al. (2020a)). However, the conventional practice of evaluating classification models via a singular evaluation metric, most often the accuracy score, does not often provide adequate information (He & Garcia (2009)). Consequently, the accuracy, recall, precision, and area under the curve (AUC) scores are each considered to evaluate the performance of the proposed ANN.
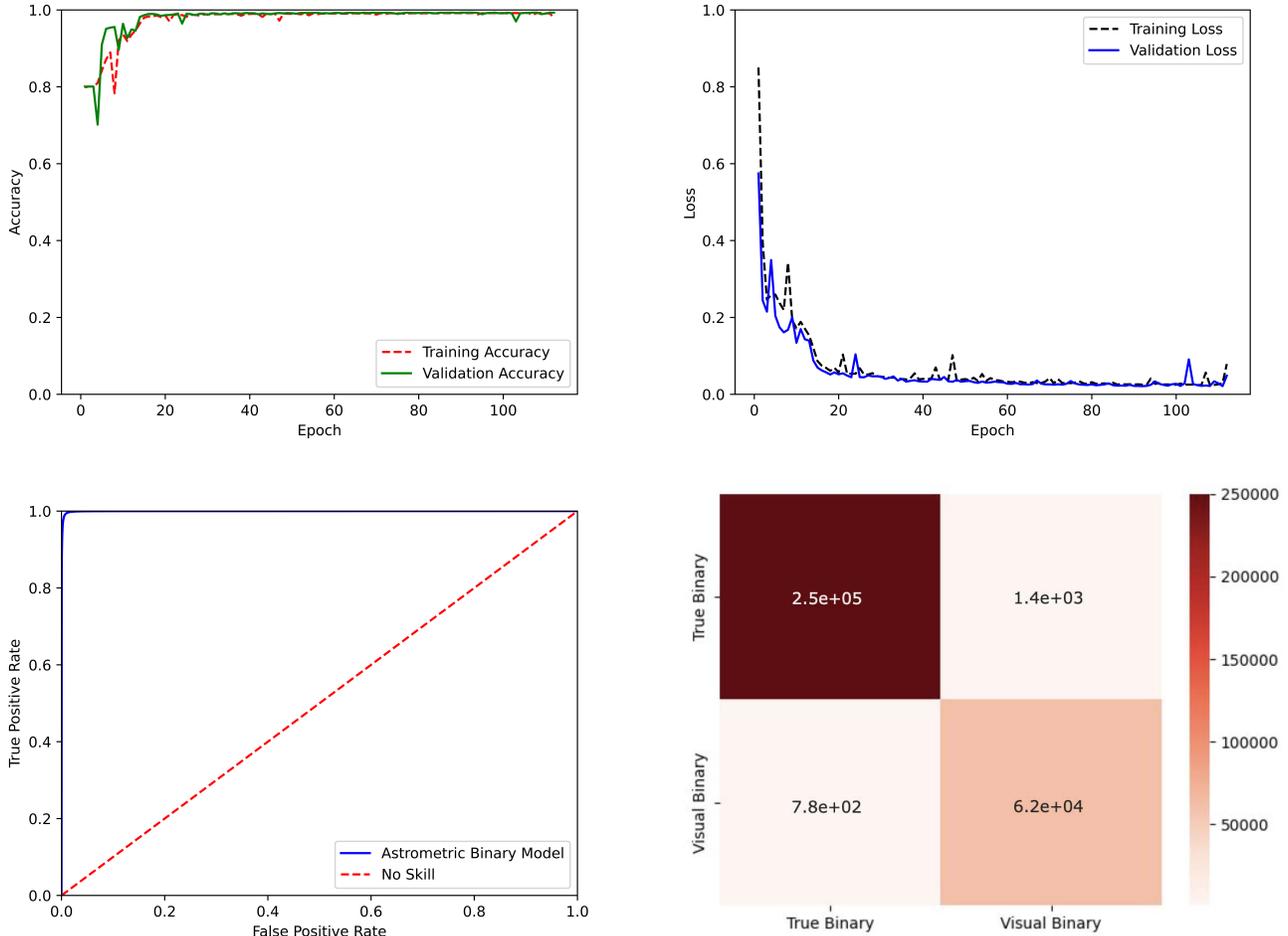
Figure 4. The first row presents the performance of the ANN considering the accuracy (left) and loss (right) evolutions during the training and validation phases. The second row presents the computed ROC curve (left) and confusion matrix (right) during the test phase.

With respect to the application of pattern recognition tasks to this work, accuracy measures how often the proposed ANN correctly classified the input binary star as either a visual binary or a true binary. Similarly, recall measures how many binaries would be correctly classified as true binaries from all the true binaries, and precision measures how many true binaries would be correctly classified from all true binary candidates. Meanwhile, the AUC score indicates how well a model can discriminate, or rank, between random examples from the data set via their computed class probability across varying thresholds. To compute an AUC score, a receiver operating characteristic (ROC) curve is necessary. The ROC curve is created by plotting the true positive rate (TPR) as a function of the false positive rate (FPR). The FPR is defined as the probability that a true binary label is given to an input visual binary while the TPR is defined as the probability that a true binary label is given

to an input true binary. When a learning algorithm has an AUC of 0.5, it is considered that the model has a random classification scheme ("no skill"). Each of these metrics have a score range of 0 to 1, with 1 representing perfect model classification.

4. RESULTS AND DISCUSSION

In the following section, the training, CV, and testing performances are each evaluated via the scikit-learn open source ML library (Pedregosa et al. (2011)). The misclassified binaries are also investigated and solutions are proposed for improving model performance. Furthermore, the influence on the data set given by the physical separation feature is investigated by constructing ML comparison models trained on a subset of the features of the proposed ANN. With current knowledge, this work is the first to develop an automatic classification tool for astrometric binary star systems.
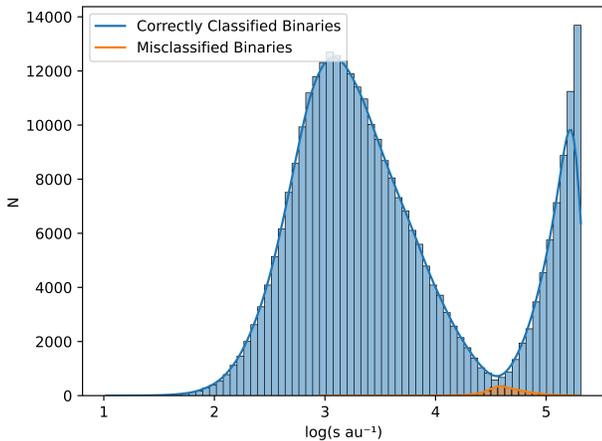
Figure 5. Histogram of misclassified and correctly classified binaries from the test set as a function of their physical separation. The range of physical separations for misclassified binaries fall between $2.95 < \log(s \text{ au}^{-1}) < 5.30$, with values falling between $4.10 < \log(s \text{ au}^{-1}) < 5.18$ within $\pm 3\sigma$ of the mean.
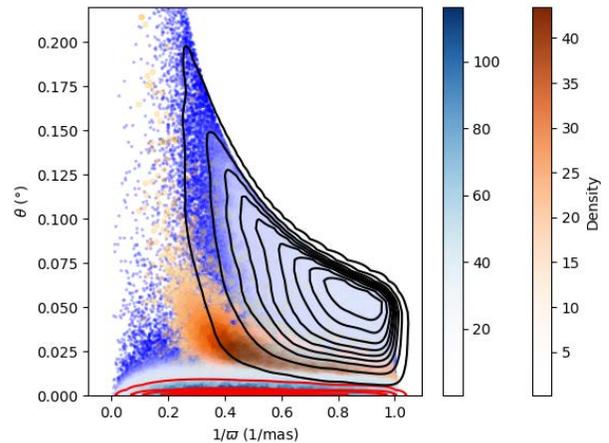

Figure 6. The angular separations of classified binaries presented as a function of the inverse of the system's primary parallax. Both sets of misclassified (orange) and correctly classified (blue) binaries are overplotted by a Gaussian kernel density estimation. The binaries in the test set are denoted by the two contours, where the black and red contours correspond to the visual and true binary classes, respectively.

4.1. *Performance and Analysis*

The evolution processes of the accuracy and loss scores as a function of epoch are shown in Figure 4. The ANN converges very quickly, reaching high accuracy and low loss scores within 30 epochs. This can be attributed to the optimization of the selection of hyperparameters and the large data set that was used to train and validate the ANN. Besides occasional deviations, the accuracy function continually increased and the loss function continually decreased before the early stopping regularization step halted the training process within 112 epochs. To evaluate model performance, a ROC curve (Hajian-Tilaki (2013)) and a confusion matrix (Visa et al. (2011)) are computed and shown in Figure 4.

From the ROC curve in Figure 4, the ANN achieves an AUC score of 0.999, indicating that the model provides incredibly high quality classifications with respect to the TPR and FPR. The high classification quality of the ANN is also shown in the confusion matrix, where nearly every input binary in the test set is correctly classified as either a true binary or a visual binary. The model achieves high classification scores, with an accuracy of 0.993, a precision of 0.988, and a recall of 0.991.

4.2. *Computational Performance*

For the training, validation, and testing process, a standard laptop computer processor unit (CPU) was utilized. The 2.6 GHz 6-Core Intel Core i7 CPU took 8 minutes to execute the entire training and validation

phases, computing 112 epochs at 4-5 s per epoch. The classification of the test set (313974 binaries) took 10 s.

4.3. *Misclassified Binaries*

In this section, the misjudged astrometric binary candidates are considered. Figure 5 presents a histogram of the misclassified binaries and the correctly classified binaries from the test set as a function of their physical separation, which is effectively a plot of the confusion matrix in Figure 4 against the physical separation feature. This histogram verifies that there is notable influence given by the physical separation feature on the dataset. Figure 6 plots the angular separations of the classified binaries against the inverse of the corresponding primary component's parallax and provides contours of the unclassified true and visual binary classes in this space. The misclassified binaries have mean angular separations of $0.021° \pm 0.01°$, a similar range of angular separations found in Figure 2 where the intersection between the two classes also occurs.

Interestingly, the majority of misclassified binaries fall under the visual binary class. This can be attributed to the dense clustering of the true binary class around the bottom edge of the model's estimated boundary between the two classes in the space found in Figure 6. Furthermore, the rate at which misclassified binaries occur follows the general rate of change of the outer boundary of the visual binary class due to the dense clustering of the

true binaries. As a result, there is a significantly greater number of false positives than false negatives.

To solve this dilemma, additional features should be considered that can provide the ANN with the information necessary to correctly classify astrometric binaries when the other parameters are unhelpful. In particular, the incorporation of the RV feature may significantly increase model performance with the release of *Gaia* DR4 (Gaia Collaboration et al. (2023)). While additional features may not always bring an increase in model performance, the selection of meaningful features, which RVs certainly are, is important for the maximization of model classification scores and can minimize the influence of the physical separation feature (Koller & Sahami (1996); Kohavi & John (1997)). Another alternative is to incorporate active learning (AL) into the modelling procedure of the ANN. AL is a technique in ML designed to optimize model performance by intelligently selecting points of interest to build a custom training data set for learning algorithms (Settles (2009)). When model performance is low in a region of classification, the AL algorithm can suggest unlabelled data that would be most helpful for the learning algorithm (Cohn et al. (1996)). The algorithm then makes queries to an oracle to determine the label for the suggested data and adds this data back into the training set to re-train the model. Considering that AL is especially useful in situations where the number of samples to label is large, an AL approach to select the relevant candidate binaries for training is a promising strategy for increasing model performance.

4.4. *Machine Learning Model Comparisons*

Given the strong definition of the true and visual binary classes by the physical separation feature, it is important to rigorously verify that the classification process is not defined by this feature. In comparison to the proposed ANN, a logistic regression (LR) model and an additional DNN, ANN-F4, are constructed. The LR model only considers the physical separation feature and ANN-F4 considers the four astrometric parameters (the proper motions and the parallaxes of the component stars) for training. Here, ANN-F4 has the same model architecture as the proposed ANN found in Figure 3. Additionally, a cutoff point in the $\log(s \text{ au}^{-1})$ space is set at $P_c = 4.61$, the mean of the overlap between the true and visual binary classes found in Figure 1, to compare the proposed ANN's classification scores at a predetermined physical separation threshold.

In Figure 7, the associated ROC curves for the P_c cut and the ANN-F4 and LR models are presented. When comparing classification models, it is usually best to consider the AUC of each model (Ling et al. (2003)). To
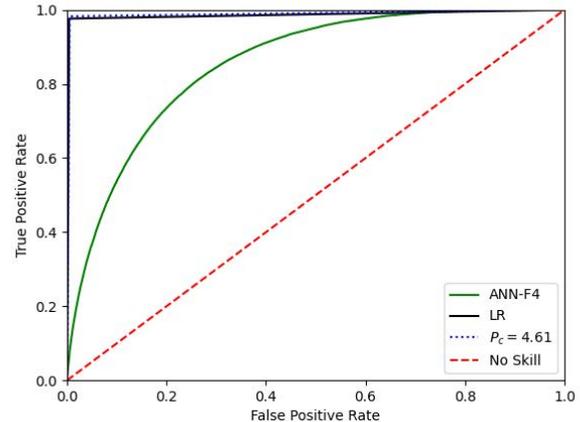
Figure 7. ROC curves for each of the comparison models are plotted.

Table 2. Statistical performance of the comparison models

Model	n	AUC	Accuracy	Precision	p-value
LR	313794	0.989	0.991	0.988	$p < 0.05$
ANN-F4	313794	0.854	0.817	0.745	$p < 0.001$
$P_c = 4.61$	1568967	0.988	0.991	0.986	$p < 0.01$

compare these comparison models to the proposed ANN, the DeLong test is utilized to provide the statistical difference between the AUCs of these models and the AUC of the proposed ANN (DeLong et al. (1988)). The DeLong test determines whether a statistically significant ($p < 0.05$) difference is present in the classification ability of the models. If a statistically significant difference is found, then it is taken that the AUC of the proposed ANN is indeed independent of the AUC of the comparison model and is thus a better classifier. The AUC, accuracy, precision, and DeLong p-value of the P_c cut and the ML models are shown in Table 2. The number of astrometric binaries in each test data set, n, are also shown in this table.

From the results shown in Table 2, these models do not perform as well as the proposed ANN. The expectation that the proposed ANN is the optimal classifier with regard to an AUC of 0.999 is verified by the DeLong test, where statistically significant tests are found for each ML model and the P_c cut. With a statistically significant test and a lower accuracy score found in the LR model and the P_c cut, it is evident that the proposed ANN is not solely defined by the physical separation feature. Instead, the additional information provided by the parallax and proper motion features permits for the proposed ANN to perform much better than these other

models. Therefore, the statistical performance of the ML comparison models and the P_c cut demonstrate that the selected features of the proposed ANN are necessary for optimal performance. It is important to note that the very statistically significant result ($p < 0.001$) found in ANN-F4 is a product of the model not being trained on the physical separation feature. Additionally, the difference in the statistical significance of the LR model and the P_c cut despite both utilizing the same feature can be attributed to the shape of each test set relative to the shape of the test set of the proposed ANN.

A final point of interest is the result shown from ANN-F4, which achieved quite high AUC and accuracy scores. While ANN-F4 clearly does not perform as well as the proposed ANN, this model does provide significant promise for the potential to classify astrometric binaries with only their observable astrometric parameters. Thus it is reasonable to suspect that the performance of ANN-F4 may increase with the widespread introduction of the RV feature that comes with the release of *Gaia* DR4.

5. CONCLUSIONS

In this paper, an artificial neural network (ANN) is proposed to classify a set of stars as either a true or visual binary based on their astrometric parameters (positions, proper motions, and parallaxes). To take advantage of the *Gaia* mission, the learning algorithm was trained on a data set comprised of an extensive number of binary samples and astrometric parameters (positions, proper motions, and parallaxes) from *Gaia* DR3. After careful testing, the ANN performs exceptionally well on withheld data, achieving very high scores in every classification metric, and is not subject to overfitting.

It was observed that the constructed data set was strongly defined by the physical separation feature. As an experiment, a logistic regression model, an additional neural network, and a predetermined cut in the physical separation space were created to compare to the proposed ANN. The comparison models were each outperformed by the ANN, implying that every selected feature is necessary for optimal model performance. In addition, this experiment demonstrated that there is significant potential for classifying astrometric binaries with regard to only their observable astrometric parameters. Therefore, it is suspected that it will become feasible to classify astrometric binaries and neglect the consideration of the physical separation and angular separation features with the robust introduction of radial velocities (RVs) when *Gaia* DR4 is published.

Compared to current classification techniques, the proposed ANN offers significant benefits for classifying astrometric binaries by saving significant net computing time and costs. Run on a cheap and low performance CPU, the training and validation phases of the ANN took about 8 minutes, while the classification of the test set (approximately 314000 binaries) took only 10 s. If the developed ANN was run on a similar CPU or graphics processing unit (GPU) compared to other machine learning (ML) models (e.g., Szklenár et al. (2020)), the above computing times would be significantly lower.

As various sky surveys continue to provide ample data sets of binary star candidates, astronomers need tools to identify astrometric binaries in an effective and efficient way. In particular, the *Gaia* mission has proven to be difficult to analyze in depth due to the sheer volume of data obtained for nearly two billion point sources. Similarly, future sky surveys that provide ample astrometric data sets require an automation tool to be fully taken advantage of. Therefore, the proposed ANN is a viable method for the analysis of large databases, thereby providing a means for astronomers to take full advantage of the *Gaia* mission and future sky surveys for other and similar research purposes.

ORCID IDS

Joe Smith https://orcid.org/0009-0008-3226-6205

REFERENCES

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/

Arlot, S., & Celisse, A. 2010, Statistics Surveys, 4, 40 , doi: 10.1214/09-SS054

Bailer-Jones, C. A. L., Irwin, M., & Von Hippel, T. 1998, Monthly Notices of the Royal Astronomical Society, 298, 361, doi: 10.1046/j.1365-8711.1998.01596.x

Banerji, M., Lahav, O., Lintott, C. J., et al. 2010, MNRAS, 406, 342, doi: 10.1111/j.1365-2966.2010.16713.x

Caruana, R., Lawrence, S., & Giles, C. 2000, in Advances in Neural Information Processing Systems, ed. T. Leen, T. Dietterich, & V. Tresp, Vol. 13 (MIT Press)

Chollet, F., et al. 2015, Keras, https://keras.io

Cohn, D. A., Ghahramani, Z., & Jordan, M. I. 1996, arXiv e-prints, cs/9603104, doi: 10.48550/arXiv.cs/9603104

DeLong, E., DeLong, D., & Clarke-Pearson, D. 1988, Biometrics, 44, 837–845, doi: https://doi.org/10.2307/2531595

Dietterich, T. 1995, ACM Computing Surveys, 27, 326, doi: 10.1145/212094.212114

Eatough, R. P., Molkenthin, N., Kramer, M., et al. 2010, Monthly Notices of the Royal Astronomical Society, 407, 2443, doi: 10.1111/j.1365-2966.2010.17082.x

El-Badry, K., & Rix, H.-W. 2018, MNRAS, 480, 4884, doi: 10.1093/mnras/sty2186

El-Badry, K., Rix, H.-W., & Heintz, T. M. 2021, MNRAS, 506, 2269, doi: 10.1093/mnras/stab323

Folkes, S. R., Lahav, O., & Maddox, S. J. 1996, Monthly Notices of the Royal Astronomical Society, 283, 651, doi: 10.1093/mnras/283.2.651

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1, doi: 10.1051/0004-6361/201629272

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1, doi: 10.1051/0004-6361/202243940

Gruner, D., Barnes, S. A., & Janes, K. A. 2023, A&A, 675, A180, doi: 10.1051/0004-6361/202346590

Hajian-Tilaki, K. 2013, Caspian journal of internal medicine, 4, 627

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. (Springer)

He, H., & Garcia, E. A. 2009, IEEE Transactions on Knowledge and Data Engineering, 21, 1263, doi: 10.1109/TKDE.2008.239

Hearst, M., Dumais, S., Osuna, E., Platt, J., & Scholkopf, B. 1998, IEEE Intelligent Systems and their Applications, 13, 18, doi: 10.1109/5254.708428

Hernandez, X., Verteletskyi, V., Nasser, L., & Aguayo-Ortiz, A. 2024, MNRAS, 528, 4720, doi: 10.1093/mnras/stad3446

Hinton, G. E., & Salakhutdinov, R. R. 2006, Science, 313, 504, doi: 10.1126/science.1127647

Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Johnson, J., & Khoshgoftaar, T. 2019, Journal of Big Data, 6, 27, doi: 10.1186/s40537-019-0192-5

Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980, doi: 10.48550/arXiv.1412.6980

Kohavi, R., & John, G. H. 1997, Artificial Intelligence, 97, 273, doi: https://doi.org/10.1016/S0004-3702(97)00043-X

Koller, D., & Sahami, M. 1996, in ICML, 284–292

Lahav, O., Naim, A., Sodré, L., J., & Storrie-Lombardi, M. C. 1996, MNRAS, 283, 207, doi: 10.1093/mnras/283.1.207

LeCun, Y., Bengio, Y., & Hinton, G. 2015, nature, 521, 436

Li, X., Zheng, Y., Wang, X., & Wang, L. 2020, The Astrophysical Journal, 891, 10, doi: 10.3847/1538-4357/ab6d04

Lin, H., Li, X., & Luo, Z. 2020a, Monthly Notices of the Royal Astronomical Society, 493, 1842, doi: 10.1093/mnras/staa218

Lin, H., Li, X., & Zeng, Q. 2020b, The Astrophysical Journal, 899, 104, doi: 10.3847/1538-4357/aba838

Ling, C. X., Huang, J., & Zhang, H. 2003, in Advances in Artificial Intelligence, ed. Y. Xiang & B. Chaib-draa (Berlin, Heidelberg: Springer Berlin Heidelberg), 329–341

Longadge, R., & Dongre, S. 2013, arXiv e-prints, arXiv:1305.1707, doi: 10.48550/arXiv.1305.1707

Medan, I., & Lépine, S. 2023, AJ, 166, 218, doi: 10.3847/1538-3881/acffb0

Nair, V., & Hinton, G. E. 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10 (Madison, WI, USA: Omnipress), 807–814

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Settles, B. 2009, Computer Sciences Technical Report 1648, University of Wisconsin-Madison

Storrie-Lombardi, M. C., Lahav, O., Sodre, L., J., & Storrie-Lombardi, L. J. 1992, MNRAS, 259, 8P, doi: 10.1093/mnras/259.1.8P

Szklenár, T., Bódi, A., Tarczay-Nehéz, D., et al. 2020, ApJL, 897, L12, doi: 10.3847/2041-8213/ab9ca4

Szklenár, T., Bódi, A., Tarczay-Nehéz, D., et al. 2022, The Astrophysical Journal, 938, 37, doi: 10.3847/1538-4357/ac8df3

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22, doi: 10.1109/MCSE.2011.37

Visa, S., Ramsay, B., Ralescu, A. L., & van der Knaap, E. 2011, in CEUR Workshop Proceedings, Vol. 710, MAICS, ed. S. Visa, A. Inoue, & A. L. Ralescu (CEUR-WS.org), 120–127. http://dblp.uni-trier.de/db/conf/maics/maics2011.html#VisaRRK11

von Hippel, T., Storrie-Lombardi, L. J., Storrie-Lombardi, M. C., & Irwin, M. J. 1994, MNRAS, 269, 97, doi: 10.1093/mnras/269.1.97

Waskom, M. 2021, The Journal of Open Source Software, 6, 3021, doi: 10.21105/joss.03021

Wes McKinney. 2010, in Proceedings of the 9th Python in Science Conference, ed. Stéfan van der Walt & Jarrod Millman, 56 – 61, doi: 10.25080/Majora-92bf1922-00a

Yoo, J., Chanamé, J., & Gould, A. 2004, ApJ, 601, 311, doi: 10.1086/380562

Zavada, P., & Píška, K. 2020, AJ, 159, 33, doi: 10.3847/1538-3881/ab5865