Received: 20 September 2022 | Accepted: 10 January 2023

DOI: 10.1111/jbg.12760

ORIGINAL ARTICLE

Journal of Animal Breeding and Genetics WILEY

Classifying aneuploidy in genotype intensity data using deep learning

Aniek C Bouwman[1] | **Ina Hulsegge**[1] | **Rachel J Hawken**[2] | **John M Henshall**[3] | **Roel F Veerkamp**[1] | **Dirkjan Schokker**[1] | **Claudia Kamphuis**[1]

[1]Wageningen University & Research, Animal Breeding and Genomics, Wageningen, The Netherlands

[2]Cobb-Vantress Inc., Siloam Springs, Arkansas, USA

[3]Cobb Vantress B.V., Boxmeer, The Netherlands

Correspondence
Aniek C Bouwman, Wageningen University & Research, Animal Breeding and Genomics, PO Box 338, Wageningen 6700AH, The Netherlands.
Email: aniek.bouwman@wur.nl

Funding information
Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin; TKI Agri & Food

Abstract

Aneuploidy is the loss or gain of one or more chromosomes. Although it is a rare phenomenon in liveborn individuals, it is observed in livestock breeding populations. These breeding populations are often routinely genotyped and the genotype intensity data from single nucleotide polymorphism (SNP) arrays can be exploited to identify aneuploidy cases. This identification is a time-consuming and costly task, because it is often performed by visual inspection of the data per chromosome, usually done in plots of the intensity data by an expert. Therefore, we wanted to explore the feasibility of automated image classification to replace (part of) the visual detection procedure for any diploid species. The aim of this study was to develop a deep learning Convolutional Neural Network (CNN) classification model based on chromosome level plots of SNP array intensity data that can classify the images into disomic, monosomic and trisomic cases. A multispecies dataset enriched for aneuploidy cases was collected containing genotype intensity data of 3321 disomic, 1759 monosomic and 164 trisomic chromosomes. The final CNN model had an accuracy of 99.9%, overall precision was 1, recall was 0.98 and the F1 score was 0.99 for classifying images from intensity data. The high precision assures that cases detected are most likely true cases, however, some trisomy cases may be missed (the recall of the class trisomic was 0.94). This supervised CNN model performed much better than an unsupervised k-means clustering, which reached an accuracy of 0.73 and had especially difficult to classify trisomic cases correctly. The developed CNN classification model provides high accuracy to classify aneuploidy cases based on images of plotted X and Y genotype intensity values. The classification model can be used as a tool for routine screening in large diploid populations that are genotyped to get a better understanding of the incidence and inheritance, and in addition, avoid anomalies in breeding candidates.

KEYWORDS

aneuploidy, B-allele frequency, chromosome, embryo transfer, SNP

1 | INTRODUCTION

Aneuploidy is the loss or gain of one or more chromosomes (e.g. monosomy or trisomy), or when all chromosomes are affected there is a change in ploidy level (e.g. haploid, triploid). These numerical chromosomal aberrations are usually lethal when occurring in autosomes, hence aneuploidy is rarely seen in living individuals, with the exception of aneuploidy in sex chromosomes and down syndrome in humans (Hassold & Hunt, 2001). In stillborn or malformed piglets (Grahofer et al., 2019), or unhatched eggs (De Boer et al., 1984; Thorne et al., 1991) aneuploidy is more prevalent. This is also the case for in vitro-produced embryos in humans and bovine (Fragouli et al., 2008; Silvestri et al., 2021; Tšuiko et al., 2017). In livestock, triploidy has been observed in adult chicken (De Boer et al., 1984; Felipe et al., 2018), as well as loss of sex chromosomes in adult sheep and cattle (Berry et al., 2017, 2018). Although a rare phenomenon in living individuals, these individuals should be identified to exclude them as breeding candidate, and to explore potential causes and heritable factors of these anomalies to reduce occurrence further. In addition, pre-screening of cattle embryos for aneuploidy plays an important role in improving pregnancy success rates of in vitro fertilization programs (Silvestri et al., 2021).

Traditionally, aneuploidy is detected by cytogenetic methods like chromosome staining, fluorescence in situ hybridization (FISH) and comparative genomic hybridization (CGH) (Martin & Warburton, 2015). Such laborious methods are in general not applied routinely on a large scale, unless individuals show fertility problems. However, routine screening is implemented for AI boars in pig breeding programs in several countries (Ducos et al., 2007, 2008). Most livestock breeding programs routinely genotype their breeding populations with SNP arrays. Therefore, SNP array data are preferred for routine aneuploidy screening of large populations. Genotype intensity data have been used to detect aneuploidy in other studies (Berry et al., 2017, 2018; Silvestri et al., 2021) and are also exploited by copy number variant calling programs like PennCNV (Wang et al., 2007).

Methods exploiting genotype intensity data are based on bioinformatics, using educated thresholds and deviations from normal samples. Those methods require visual inspection of the data from detected cases for confirmation, as false positives are common.

Since visual inspection of plots from genotype intensity data by experts is, thus, the most commonly applied method to verify aneuploidy based on SNP array data, we were interested in whether image classification models could automate the visual inspection of images. Moreover, image classification is also interesting because the visual inspection is based on images that are created and assessed on chromosome level. This makes image classification generic across diploid species and SNP arrays, because it does not matter how many chromosomes a species has or how many SNPs on a chromosome.

Deep learning algorithms like Convolutional Neural Networks (CNNs) are very powerful in image recognition (Krizhevsky et al., 2012). CNNs are especially equipped for recognizing patterns. This makes CNNs an interesting tool to recognize the clustering patterns of genotype intensity data into three clusters for disomic chromosomes (AA, AB and BB), two clusters for monosomic chromosomes (A, B) and four clusters for trisomic chromosomes (AAA, AAB, ABB and BBB). CNN can exploit the spatial relationship between all datapoints in the image. A requirement to build an accurate CNN Classification model, however, is sufficient training data. This requirement may be problematic due to the low incidence of aneuploidy and data availability in a single species. Hence, genotype intensity data cases collected across species, especially in datasets expected to be enriched for aneuploidy cases (e.g. unhatched eggs, in vitro produced embryos, fish), seems an attractive approach to generate a dataset of sufficient size to develop a model that is generic across diploid species.

Therefore, the aim of this study was to develop a deep learning CNN classification model based on chromosome level plots of SNP array intensity data that can be used as a tool for routine screening for aneuploidy in large diploid populations genotyped with SNP arrays.

2 | MATERIALS AND METHODS

2.1 | Datasets

Genotype intensity data were retrieved from three different datasets. (a) A chicken dataset (Cobb Vantress) enriched for chromosomal anomalies, containing 2647 chickens genotyped with a 60 K custom Infinium SNP array (Illumina, Inc.). (b) A cattle embryo dataset (CRV B.V.) of 120 genotyped embryos enriched for chromosomal anomalies. The embryos were genotyped with either the 10 K Eurogenomics SNP array or the 50 K Eurogenomics SNP array (Illumina Inc.). (c) A fish dataset (Hendrix Genetics) with 472 salmons that were genotyped on a 25 K custom multispecies SNP array (Illumina Inc.). Salmon is a species that can occur as triploid, in this dataset two triploids were detected. Furthermore, the multispecies chip contained a set of salmon-specific SNP and a set of trout-specific SNP. The salmons appeared as if they were haploid for the trout SNP set. This created artificially a large pool of monosomy cases to ensure sufficient annotated data for the purpose of this study.

In addition, SNP marker map files were available for each SNP array, indicating the position of the SNP on the genome, including chromosome and location on chromosome in base pairs.

2.2 | Genotype intensity data annotation

In all three datasets, the genotype intensity data contained at least normalized X and Y intensity levels. Normalized X and Y intensity values represent the colour intensity of the fluorescent green and red dye representing the two alleles (A and B) of the SNP. Aneuploidy can be detected by plotting these normalized X and Y intensity values in a scatter plot, hereafter referred to as XY intensity plot. For segregating SNPs in diploid individuals there are usually three clusters in XY intensity plots, in which the X intensity values were plotted along the X axis and the Y intensity values were plotted along the Y axis (Figure 1A). These three clusters represent three genotypes: one cluster along the X axis with Y intensity around zero, representing AA genotypes. A second cluster along the Y axis with X intensity around zero, representing BB genotypes. And a third cluster along the diagonal representing the heterozygous AB genotypes. In case of monosomy, the heterozygous genotypes (AB) are absent, showing only two clusters, one along the X axis for genotype A and the other along the Y axis for genotype B (Figure 1B). In case of trisomy, the XY intensity plots have four clusters, one for each genotype: two homozygous clusters along the two axis for AAA (X axis) and BBB (Y axis) genotypes; and two heterozygous clusters, AAB below the diagonal and ABB above the diagonal (Figure 1C).

Each individual was screened for aneuploidy by visual inspection of the XY intensity plots and additional genotype intensity information when available (e.g. B-allele frequency (BAF) and LogR Ratio (LRR)) by one of two experts. LogR Ratio, is the $\log_2$ of the normalized R value (R = X+Y) over the expected R value. The R expected is based on a set of diploid genotyped individuals representative for the respective specie and breed. Each chromosome was evaluated separately. The sex chromosomes were excluded. Working on chromosome level meant that haploid or triploid cases counted as monosomic or trisomic, respectively, for all their autosomes. This led to a focus on three classes (Figure 1): (a) disomic; (b) monosomic and (c) trisomic.

2.3 | Model development data

After annotation, images of XY intensity plots were created for each chromosome of each annotated individual as input data for CNN model development. The initial annotated dataset was highly unbalanced, with mainly disomic examples, and limited trisomic examples. Hence, a subset of the disomic examples was randomly selected, while all monosomic and trisomic examples were kept. The dataset was still unbalanced, with almost twice the number of disomic cases compared to monosomic cases, but disomic cases will also be the most abundant and variable group in real data. The remaining images were once again visually inspected to remove spurious cases.

This cleaned annotated dataset was randomly split into a training set containing 60% of the data per class, a validation set containing 20% of the data per class and a test set containing 20% of the data per class (Table 1). The split was based on chromosomal plots, not on individuals, hence different chromosome plots of the same individual can be split into different sets. Images for training and validation were used for model development and internal fine-tuning of model parameters. Images in the test set were, thus, independent (on chromosome level) from the training model and used for model evaluation on unseen data.

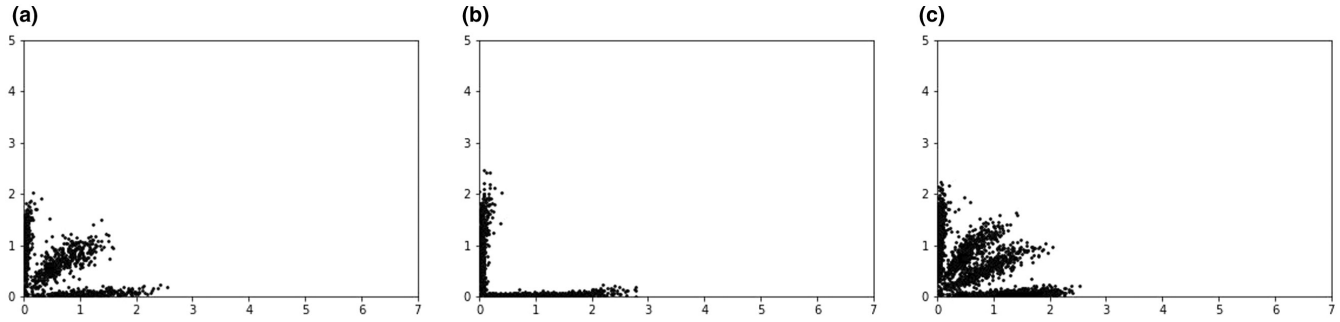

FIGURE 1 Example images of chromosomal XY intensity plot of class disomic (a), monosomic (b) and trisomic (c). The normalized intensity levels of the B-allele are on the X-axis and the normalized intensity levels of the A-allele on the Y-axis, each dot represents a SNP mapped to the chromosome.

2.4 | Baseline model with unsupervised k-means clustering

Classification can be done both supervised and unsupervised. An unsupervised classification does not require annotation of data, which saves a lot of time. However, unsupervised model may be less accurate if the task is too complex. An unsupervised k-means clustering algorithm was applied to the data as a baseline, to check its performance, and be able to assess the improvement when using supervised learning. A k-means clustering algorithm was trained in python using the training dataset, and tested on the test dataset described in Table 1. For training, 3145 chromosomal XY intensity plots were read with a size of 224 pixels × 224 pixels × 3 (RGB colour code) and flattened from a 3D array to a 1D array of 150,528 columns. Dimension reduction was needed, hence a PCA was performed on the 3145 images used for training. The first three components explained 27.96%, 7.21% and 3.85% of the variation, with the first and second component differentiating the monosomic cases from the rest (Figure S1). K-means clustering was performed on all 3145 PC's. We varied the number of clusters (k) from 1 to 10, where each k-means algorithm was run 10 times with different centroid seeds, and a maximum of 300 iterations for a single run. The Within-Cluster Sum of Square (WCSS) is the sum of squared distance between each point and the centroid in a cluster. The WCSS was plotted for each number of clusters in an elbow plot. The changing point creating the elbow shape in the plot is called the elbow (or knee) point and indicates the most optimal number of clusters for the training dataset, this was estimated using the KneeLocator function in python. The most optimal k-means algorithm was tested on the 1050 chromosomal XY intensity plots from the test dataset.

TABLE 1 Amount of annotated chromosomes per class available for model development.

	Number of chromosomes			
Class	**Total**[a]	**Training**	**Validation**	**Test**
Disomic	3,321	1,992	664	665
Monosomic	1,759	1,055	352	352
Trisomic	164	98	33	33
Total	5,244	3,145	1,049	1,050

[a]Total number of each class after cleaning step.

2.5 | Supervised Convolutional Neural Network classification model

For the classification task, a Convolutional Neural Network (CNN) was developed. For the CNN, we used a VGG-like architecture (Simonyan & Zisserman, 2015), consisting of two major layers: a feature extraction layer and a classification layer (Figure 2). The feature extraction layer was composed of four convolution layers where each layer is followed by a pooling layer. Once the features were extracted, they were mapped by a subset of fully connected layers to the final output of the network, being the probabilities for each class in the classification task. The final fully connected layer has the same number of output nodes as the number of classes (disomic, monosomic, trisomic).

The model was built using Keras (version 2.3.1) (Chollet, 2018) with a TensorFlow backend (version 2.0.0; Martin Abadi et al., 2016) in Python (version 3.6). The CNN was trained for 25 epochs.

2.6 | Model performance assessment

The performance of both the unsupervised baseline k-means clustering model and the supervised CNN clustering model was assessed using a confusing matrix and model performance parameters. The confusion matrix of the test set shows the combination of the actual and predicted classes. It provided insight in which classes were easily confused by the classification model. For the CNN model, wrongly classified examples in the test set were inspected visually to judge if these were complex examples. Also, the probabilities of the prediction for each class of the wrongly classified examples were investigated.

The model assessment parameters precision (true positives/[true positives + false positives]), recall (true positives/

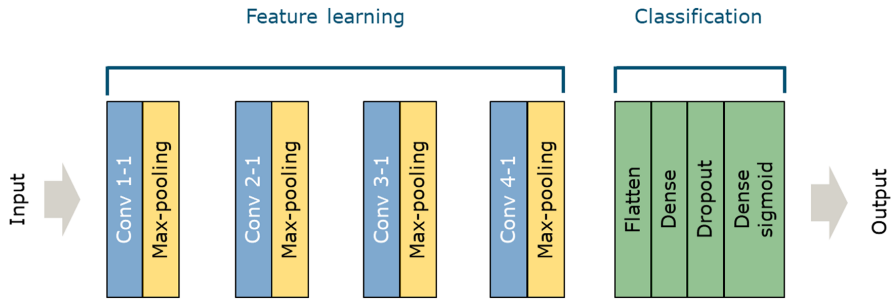

FIGURE 2 The schematic of the architecture of VGG-like CNN. [Colour figure can be viewed at wileyonlinelibrary.com]

[true positives + false negatives]) and F1 score (2*((precision × recall)/(precision + recall))) were assessed on a per class basis, and at macro and micro-average level for evaluating performance average across classes. A macro-average computes the performance metric independently for each class and then takes the global mean, hence treating all classes equally. Whereas, a micro-average aggregates the contributions of all classes to compute the average metric, hence the metric is weighted by the number of samples from each class. In a multi-class classification setup, micro-average is preferable if the classes are unbalanced. (Yang, 1999).

3 | RESULTS

The clustering models were developed using a multi-species dataset consisting of images from 3321 disomic chromosomes, 1759 monosomic chromosomes and 164 trisomic chromosomes annotated by experienced researchers (Table 1). Figure 1 shows an example image for each class.

3.1 | Plot settings

Plots can be created in many different ways and we tested a number of plot settings to assess if the CNN model would perform differently. The plot settings tested were as follows: (a) size of points; (b) open or filled points; (c) axis scale fixed or flexible, or a combination thereof (Figure 3). The results showed limited differences (Table 2), but small filled dots with fixed axis (Figure 3B) performed best in terms of accuracy for the test set. Making the shape smaller (Figure 3B,D) or open (Figure 3C,D) increased performance slightly compared to the original plots (Figure 3A), while flexible axis scale (Figure 3E) reduced performance (Table 2).

3.2 | Baseline model with unsupervised k-means clustering

An unsupervised k-means clustering algorithm was used as a baseline model. We varied the number of clusters (K) from 1 to 10 and plotted the Within-Cluster Sum of Square (WCSS; Figure S2), the elbow point indicated that three clusters are most optimal for the training dataset. Testing the k-means clustering algorithm with three clusters on an unseen test dataset (unseen chromosomes, not independent in terms of individuals) resulted in an accuracy of 0.73. The confusion matrix in Table 3 and model performance in Table 4 show that monosomic examples can be classified rather well with an error rate of 0.003, and recall of 1, although the precision is 0.80. Disomic examples have an error rate of 0.376, and reasonably high precision (0.97), but low recall (0.62). While trisomic examples are very problematic with an error rate of 0.879, low precision (0.02), recall (0.12) and hence a low F1 score of 0.04. These results suggest a model is needed that is better at classifying trisomic and disomic cases.

3.3 | Supervised Convolutional Neural Network classification model

The final CNN classification model was based on XY intensity plots with small closed points and fixed axis like Figure 2B. Tables 5 and 6 show the results of the model performance on an unseen test dataset (unseen chromosomes, not independent in terms of individuals). The confusion matrix in Table 5 indicates that the class trisomic was most difficult to classify correctly with two false negatives out of 33 examples. This is, however, much better than the unsupervised k-means clustering performance on the trisomic class. For the classes disomic and monosomic the precision, recall and F1 score were all 1, for trisomic the precision was 1, whereas recall was 0.94 and, therefore, the F1 score was 0.97 (Table 6). The high micro- and macro-averages of the performance parameters (Table 6) indicated that the data unbalance did not affect the model performance parameters.

The training accuracy of the CNN remained consistent after eight epochs at 99.9% and the CNN showed a consistent validation accuracy after 12 epochs at 99.4% (Figure 4A). The training loss remained rather consistent after eight epochs around 0.004, while the validation loss varied slightly around 0.025 after four epochs (Figure 4B). Hence, we were able to develop a model with high accuracy levels within a small number of training epochs that did not show signs of overfitting.

3.4 | Misclassified examples

Two trisomic examples were not correctly classified by the CNN model, and were visually inspected (Figure 5). The trisomic case in Figure 5A was predicted as disomic with a probability of nearly 1 (probability trisomic: 3.50 E-13; probability monosomic: 1.97 E-15). The two heterozygous clusters expected for a trisomic case were not clearly differentiated, rather one big cluster in the middle as is expected for disomic class, which could explain the confidence of the model. The visual evidence from XY intensity plot, the BAF plotted against genome position and from the LRR plotted against genome position for this

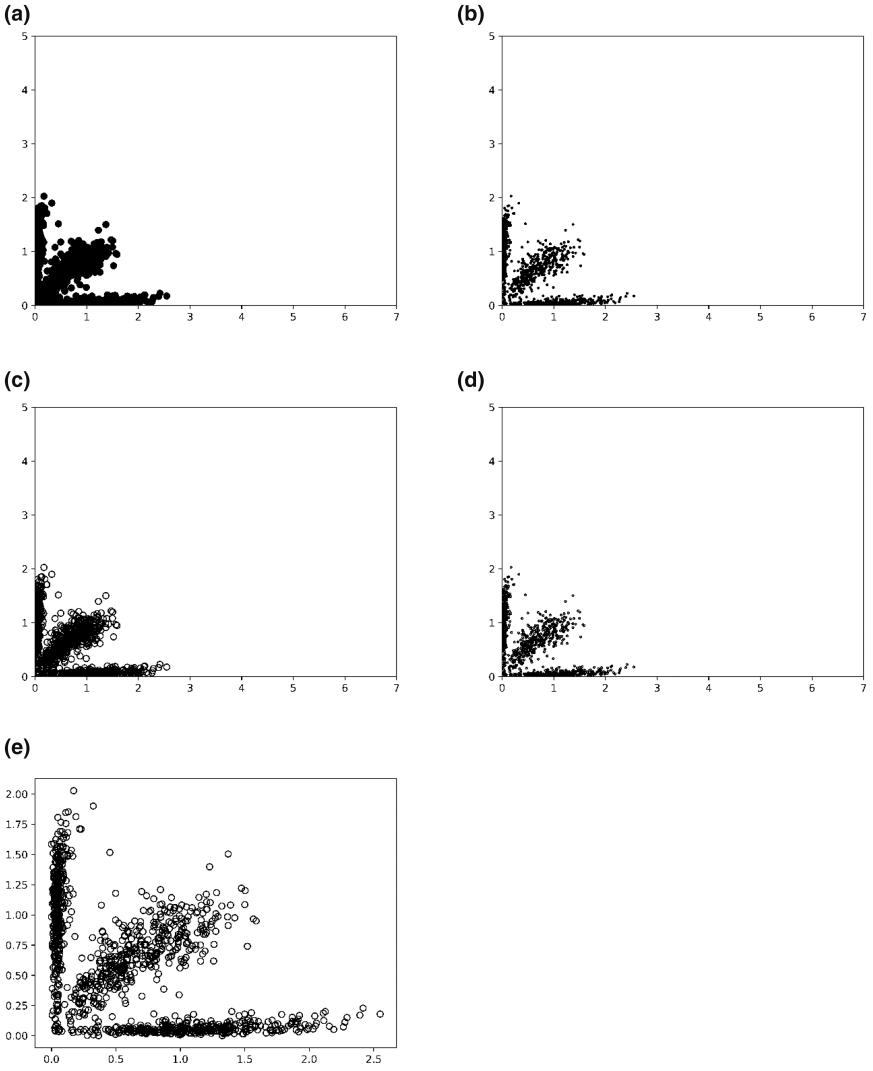

FIGURE 3 Different plot settings tested compared to the original plots (a). Smaller points (b), open points (c), a combination of smaller and open points (d), and open points with flexible axis (e) were tested.

TABLE 2 Results CNN classification models using different plot features.

	Training		Validation		Test	
Plot setting	**Accuracy**	**Loss**	**Accuracy**	**Loss**	**Accuracy**	**Loss**
Original	0.9984	0.0073	0.9946	0.0195	0.9923	0.0646
Small	0.9969	0.0172	0.9848	0.0665	0.9952	0.0479
Open	0.9987	0.0054	0.9948	0.0238	0.9942	0.0524
Open + small	0.9973	0.0112	0.9928	0.0295	0.9942	0.0484
Open + flex[a]	0.9957	0.0145	0.9896	0.0342	0.9845	0.0748

[a]flex: flexible axis range depending on X and Y values, rather than fixing y-axis at 5 and x-axis at 7.

TABLE 3 Confusion matrix unsupervised k-means clustering algorithm.

		Predicted			
		Disomic	**Monosomic**	**Trisomic**	**Error**
Annotated	Disomic	415	73	177	250/665 = 0.376
	Monosomic	0	351	1	1/352 = 0.003
	Trisomic	12	17	4	29/33 = 0.879

chromosome being trisomic is also low. However, the individual was classified as triploid, as a clearer trisomic pattern was observed on the other chromosomes of this individual.

The trisomic case in Figure 5B was predicted as monosomic with a probability of 0.99 (probability trisomic 0.005; probability disomic 0.0007). The heterozygous clusters were close to their nearest homozygous counterpart (i.e. AAB to AAA), in addition, the number of heterozygous SNPs was limited. The BAF and LRR plotted against the genome position showed long stretches of homozygosity across the chromosome due to inbreeding (LRR consistent across the chromosome), with a small region of trisomy, hence the small amount of heterozygous SNP (Figure S3). In these kinds of examples, human scorers have the advantage to be able to compare to the other chromosomes of the individual before classification, as this was clearly a triploid case from a highly inbred line (Figure S3). This does indicate that high inbreeding may cause false (monosomy) classifications, that BAF and LRR plotted against the genome position can provide more information, and that all chromosomes should be evaluated to give final conclusion on underlying ploidy issue.

Both misclassified trisomic cases were hard to classify based on X and Y intensity information from one chromosome only. The trisomic class is also for trained experts the most complicated class of these three classes to classify.

3.5 | Aneuploidy screening tool

The CNN classification model for aneuploidy is incorporated in a software tool to screen batches of genotype intensity data for aneuploidy on a routine basis as soon as new batches of animals are genotyped. The tool can take input files from both Illumina and Affymetrix genotype arrays. It outputs a text file listing the predicted class for each analysed chromosome of each individual, and provides the images of the ones classified as monosomic and trisomic for further investigation. Filtering on sample call rates and excluding specific chromosomes (e.g. exclude Y and MT) can be done. The SNP density per chromosome can vary, but a minimum number of SNP per chromosome is required. The generic workflow of this aneuploidy screening tool is given in Figure 6.

TABLE 4 Model performance unsupervised k-means clustering algorithm.

	Precision	Recall	F1 score	Support
Disomic	0.97	0.62	0.76	665
Monosomic	0.80	1.00	0.89	352
Trisomic	0.02	0.12	0.04	33
Macro Average	0.60	0.58	0.56	1,050
Micro Average	0.88	0.73	0.78	1,050

3.6 | Computation time

Computation time of different aspects of the aneuploidy screening tool was assessed on a Linux High Performance Cluster (x86_64 GNU/Linux), with an Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz processor. The results are presented in Table 7. The aspects of the aneuploidy screening tool that are independent of the sample size are "importation of libraries," "reading of instruction file," and "creation of directories." It only took 5 s all together for these three aspects to run. The aspects "read final report, map file, call rates" and "generate and store plots" are dependent of the number of individuals, number of markers on the SNP array, and the latter also on the number of chromosomes of the species. The aspects "predict class" and "remove temporary directory" are dependent of the number of individuals and number of chromosomes of the species. The number of chromosomes varies across species, and SNP array have varying densities in terms of number of SNPs. For this computational test, we used a dataset of 48 individuals with 20,502 SNP, and used 29 chromosomes, leading to 1392 plots to be generated and classified. The total computation time was 4 m 46 s. The plots were temporarily stored and hence this process had the longest duration.

4 | DISCUSSION

The aim of this study was to develop a deep learning CNN classification model based on chromosome level plots of SNP array intensity data and incorporate it in a tool for

TABLE 5 Confusion matrix.

		Predicted			
		Disomic	Monosomic	Trisomic	Error
Annotated	Disomic	665	0	0	0/665 = 0.00
	Monosomic	0	352	0	0/352 = 0.00
	Trisomic	1	1	31	2/33 = 0.06

routine screening for aneuploidy in large genotyped populations that is general across diploid species. The chromosome level approach had the advantage that we could combine multiple diploid species into one annotated dataset. This resulted in a unique annotated dataset with sufficient cases for model development, even though cases tend to be rare. Working on chromosome level also means that the resulting model can predict aneuploidy for any diploid species, regardless of the number of chromosomes or density of SNP array.

High precision of the CNN model is important for routine screening of large genotyped populations where selection decisions need to be taken early in life. With the achieved high precision levels, monosomic and trisomic cases automatically identified by the aneuploidy screening tool can be excluded from the genotyped population without any human validation or inspection. However, a small proportion of true cases may pass as disomic and go unnoticed, especially trisomic cases classified as disomic, which was shown by this study (Figure 5A). Trisomy is the most difficult class to discriminate based on XY intensity plots. Moreover, this class had the least number of examples in the dataset, resulting in an unbalanced dataset for model development. These two factors may lead to relatively more false classifications of actual trisomic cases, than for disomic and monosomic cases.

Since aneuploidy is a rare event (Hassold & Hunt, 2001), it is a challenge to obtain sufficient samples of cases. Data from unhatched eggs and cattle embryos contributed to the amount of cases, but also data from liveborn triploid chicken and salmon. Even though we combined data from different species to obtain sufficient data of cases, the number of trisomy cases in our dataset was limited, but still sufficient to obtain a good model performance. A balanced model with equal number of disomic and monosomic cases as we had trisomic cases available (e.g. 98 for training, 33 for validation and 33 for testing) showed an accuracy of 0.98 (results not shown), only misclassifying trisomic cases. Given that the unsupervised model

TABLE 6 Model performance.

	Precision	Recall	F1 score	Support
Disomic	1.00	1.00	1.00	665
Monosomic	1.00	1.00	1.00	352
Trisomic	1.00	0.94	0.97	33
Macro average	1.00	0.98	0.99	1,050
Micro average	1.00	1.00	1.00	1,050

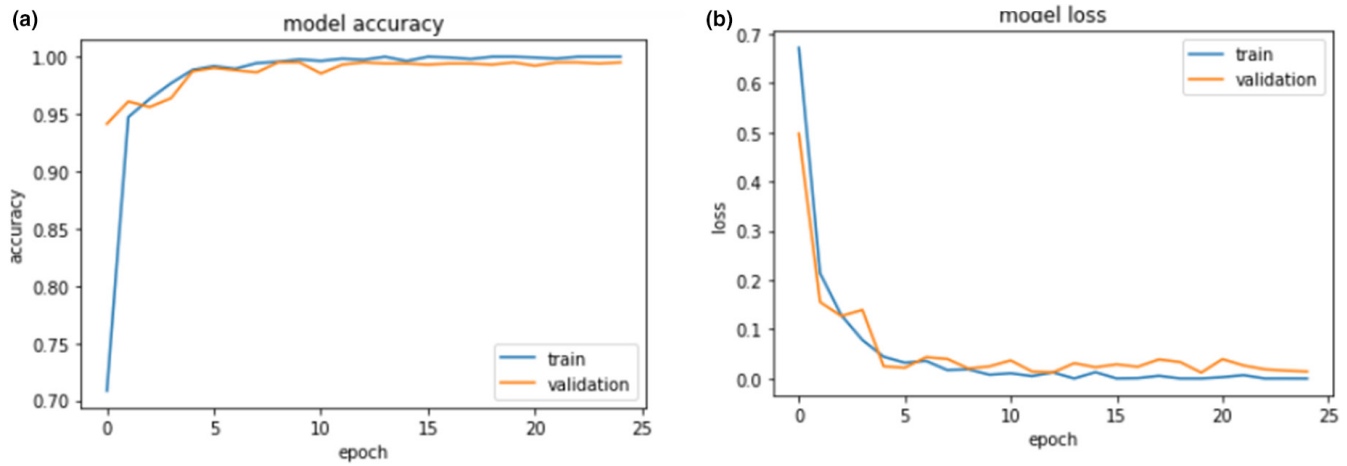

FIGURE 4 Accuracy (a) and loss (b) precision curves for the CNN classification model. The blue line denotes the training data and the orange line the validation data. The number of epochs is on the x-axis. [Colour figure can be viewed at wileyonlinelibrary.com]

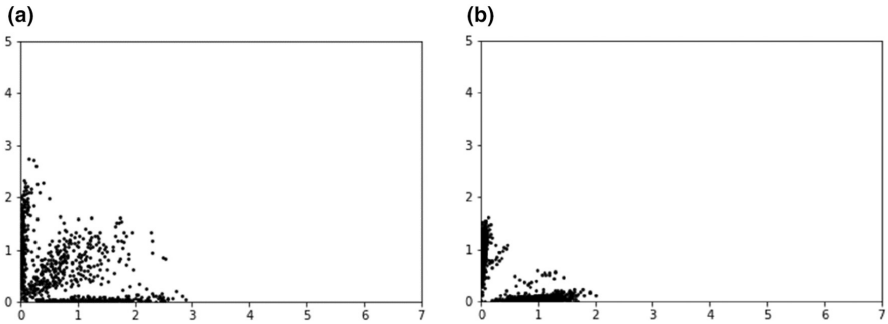

FIGURE 5 XY intensity plots for misclassified trisomic examples, (a) was classified as disomic, and (b) as monosomic.

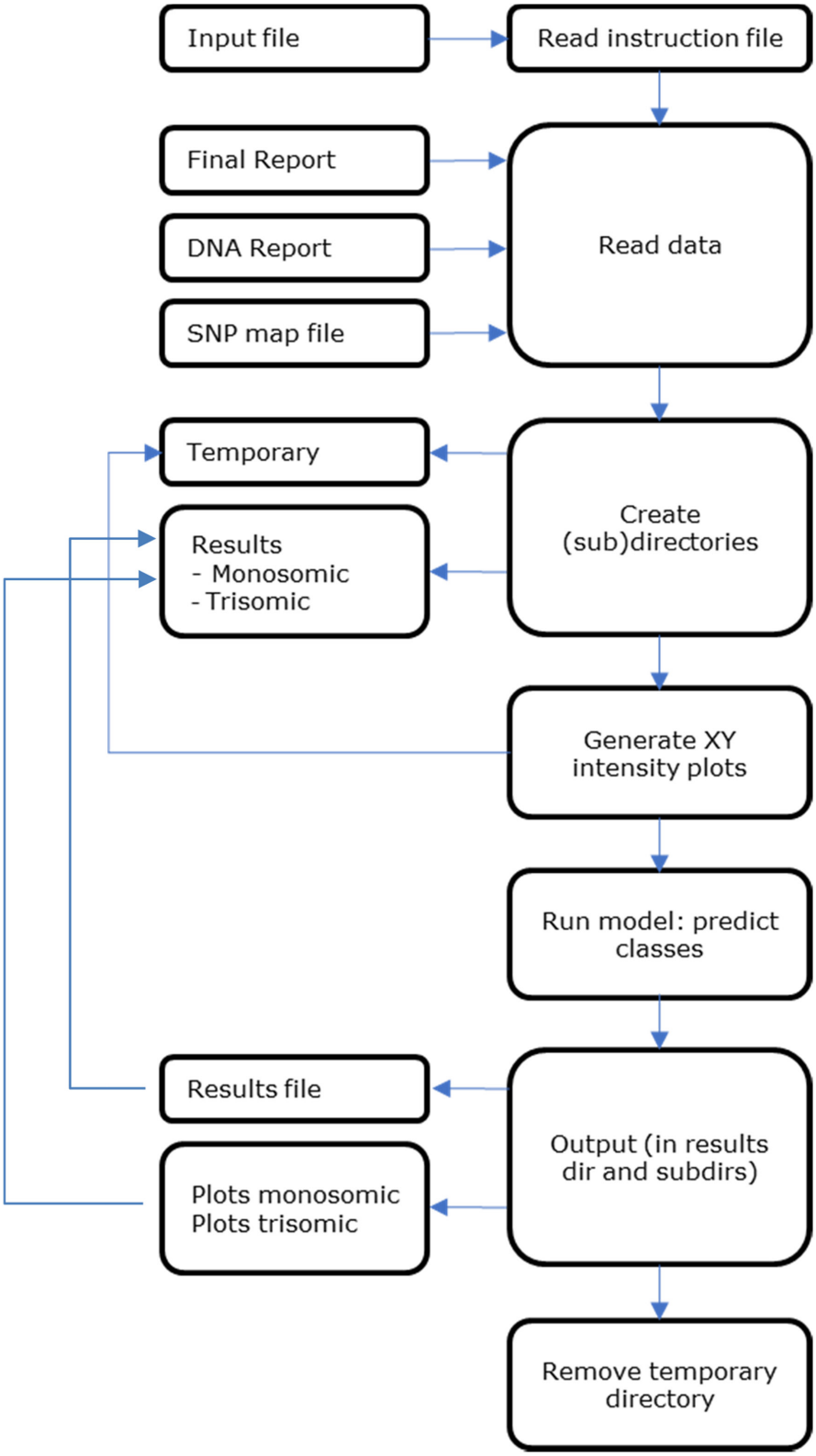

FIGURE 6 Aneuploidy screen tool workflow. [Colour figure can be viewed at wileyonlinelibrary.com]

performed well on monosomic cases, their training data need not be extremely high. However, more trisomic examples could be beneficial for model performance, as this is the most difficult class to classify, also by human experts. The monosomic class was enlarged by exploiting the multi-species SNP array in the fish data in the current study. They are not true monosomic cases, but they appear monosomic for the chromosomes of the other fish species. Augmentation of trisomy cases could potentially be done in the future by incorporating genotype intensity

TABLE 7 Computation time of different parts of aneuploidy screening tool in seconds (s).

	Total (s)	Average per ID-chr (s)
1. Import libraries	4.59	
2. Read instruction file	0	
3. Read final report, map file, call rates	6.80	0.005
4. Create directories	0.01	
5. Generate and store plots	251.94	0.181
6. Predict class	20.70	0.015
7. Remove temporary directory	1.09	

data from triploid species (e.g. in plants). The annotated dataset can be further expanded over time by adding novel annotated cases to the training data. Alternatively, cases could be simulated to augment training data or by manipulating the data from existing cases, for example, by applying label-preserving transformations to the images or underlying data. This is called artificial data augmentation and is accepted in neural network training to increase the training dataset to improve model performance, especially when overfitting occurs due to insufficient training data (Shorten & Khoshgoftaar, 2019). However, there were no signs of overfitting based on the classification accuracy plotted over epochs, suggesting that our developed CNN model generalizes well to unseen data.

The genotype intensity data of our study were used to create images for the CNN classification model, but the data can also be used directly in bioinformatic approaches, like Berry et al. (2017, 2018). Berry et al. (2017, 2018) exploited genotype intensity data, to screen large genotyped livestock populations, by identifying extreme deviations from mean intensity values across the population and LRR within individuals. They detected cases of monosomy of the X chromosome in female cattle and sheep. Setting thresholds on deviation from a population mean can be a challenge (Berry et al., 2017). Whether this method is also capable of detection trisomy, haploidy and triploidy remains unknown as it was (most likely) not present in the healthy population screened.

Most bioinformatic methods usually require visual inspection of the data from detected cases for confirmation, as false positives are commonly expected. The current study was inspired by this visual inspection. Hence, the desire to use plots of genotype intensity data directly, rather than the underlying raw data. The novelty here is to use CNN to classify XY intensity plots. CNN are known for their power to classify images using purely supervised learning and more and more used in the field of genetics. Kern and Schrider (Kern & Schrider, 2018) used images from summary statistics of population genetic features to classify neutral regions, soft sweeps and hard sweeps. Also, CNN-based variant calling methods using images of sequence reads rather than bioinformatics perform well (Poplin et al., 2018). All these CNN models were trained without specialized knowledge about genomics or the underlying data, and yet learned to classify the required classes accurately.

Differentiation between technical genotyping issues and true abnormalities remains difficult. Technical genotyping issues are expected to be random or affect the full genome, but certainly not a single chromosome, as SNPs are placed on arrays at random. Parental genotypes could be used to identify parent of origin for the missing or extra chromosome, as a confirmation of true aneuploidy (Johnson et al., 2010; Sills et al., 2014; Silvestri et al., 2021).

For liveborn individuals, sex-chromosome aneuploidy might be very relevant, as this may be one of the few tolerated aneuploidies (Berry et al., 2017, 2018; Bugno-Poniewierska & Raudsepp, 2021). Sex-chromosome aneuploidy may lead to fertility issues (Berry et al., 2017; Bugno-Poniewierska & Raudsepp, 2021) which will become apparent at reproductive age, so early detection will avoid economic losses to breeders. The aneuploidy screening tool currently does not take into account the (presumed) sex of the individual. Hence, it is an important aspect to properly include into the aneuploidy screening tool in the future.

In addition, the CNN model cannot discriminate between monosomy and a run of homozygosity (ROH) spanning a full chromosome. Only LRR information can be useful to discriminate monosomy from ROH. It was, however, decided not to include LRR information because LRR appears normal for triploids and haploids, which we also wanted to detect. Cases of (nearly) full chromosomal ROH are hence classified as monosomy and need to be further checked using LRR levels or parental genotypes. They may be considered as false positives, but it is also worthwhile to monitor such cases in breeding programs or at least be aware of it as it may indicate high inbreeding levels. This does mean the tool is not well suited for highly inbred populations.

5 | CONCLUSIONS

The developed CNN classification model provides high accuracy to classify aneuploidy cases based on images of plotted X and Y intensity values. The classification model is incorporated into an automated aneuploidy screening tool useful for routine screening in large diploid populations to avoid anomalies in breeding candidates. Routine screening for aneuploidy can henceforth be added to the utilization of genotypes from SNP arrays without additional costs other than computational costs.

FUNDING INFORMATION

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food project 16,022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin. The use of the HPC cluster has been made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR).

CONFLICT OF INTEREST STATEMENT

Co-authors John M Henshall and Rachel J Hawken are employed by Cobb Vantress, this study was co-financed by Cobb Vantress.

DATA AVAILABILITY STAEMENT

The chicken genotype intensity data that support the findings of this study were made available under licence from Cobb Vantress. The cattle genotype intensity data that support the findings of this study were made available under licence from CRV B.V. The fish genotype intensity data that support the findings of this study were available under licence from Hendrix Genetics. Restrictions apply to the availability of these data. Data may be available from the authors upon reasonable request and with permission of Cobb Vantress, CRV B.V. and Hendrix Genetics.

ETHICAL APPROVAL

The genotype data used in this study were pre-existing data generated during routine practice of the breeding companies providing the genotype data.

ORCID

Aniek C Bouwman https://orcid.org/0000-0001-5079-7108
Ina Hulsegge https://orcid.org/0000-0002-7015-6540
Roel F Veerkamp https://orcid.org/0000-0002-5240-6534
Dirkjan Schokker https://orcid.org/0000-0001-6082-7227
Claudia Kamphuis https://orcid.org/0000-0001-5552-036X

REFERENCES

Berry DP, O'Brien A, O'Donovan J, McHugh N, Wall E, Randles S, McDermott K, O'Connor RE, Patil MA, Ho J, Kennedy A, Byrne N, Purfield DC. Aneuploidy in dizygotic twin sheep detected using genome-wide single nucleotide polymorphism data from two commonly used commercial vendors. *Animal*. 2018;*12*:2462–2469.

Berry, D. P., Wolfe, A., O'Donovan, J., Byrne, N., Sayers, R. G., Dodds, K. G., McEwan, J. C., O'Connor, R. E., McClure, M., & Purfield, D. C. (2017). Characterization of an x-chromosomal non-mosaic monosomy (59, x0) dairy heifer detected using routinely available single nucleotide polymorphism genotype data1. *Journal of Animal Science*, *95*, 1042–1049.

Bugno-Poniewierska, M., & Raudsepp, T. (2021). Horse clinical cytogenetics: Recurrent themes and novel findings. *Animals*, *11*, 831.

Chollet, F. (2018). *Keras: The python deep learning library*, Astrophysics Source Code Library.

De Boer, L. E. M., De Groen, T. A. G., Frankenhuis, M. T., Zonneveld, A. J., Sallevelt, J., & Belterman, R. H. R. (1984). Triploidy in gallus domesticus embryos, hatchlings and adult intersex chickens. *Genetica*, *65*, 83–87.

Ducos, A., Berland, H.-M., Bonnet, N., Calgaro, A., Billoux, S., Mary, N., Garnier-Bonnet, A., Darré, R., & Pinton, A. (2007). Chromosomal control of pig populations in france: 2002–2006 survey. *Genetics Selection Evolution*, *39*, 583.

Ducos, A., Revay, T., Kovacs, A., Hidas, A., Pinton, A., Bonnet-Garnier, A., Molteni, L., Slota, E., Switonski, M., Arruga, M. V., van Haeringen, W. A., Nicolae, I., Chaves, R., Guedes-Pinto, H., Andersson, M., & Iannuzzi, L. (2008). Cytogenetic screening of livestock populations in europe: An overview. *Cytogenetic and Genome Research*, *120*, 26–41.

Felipe, V. P. S., Okimoto, R., Cleveland, B., Freitas, M., Grose, J., Larson, S., Katanbaf, M., Henshall, J., de Hollander, C., & Hawken, R. (2018). *Triploidy in broiler chickens: A brief review and case description*. Proceedings of the World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand, pp. 715.

Fragouli, E., Lenzi, M., Ross, R., Katz-Jaffe, M., Schoolcraft, W. B., & Wells, D. (2008). Comprehensive molecular cytogenetic analysis of the human blastocyst stage. *Human reproduction*, *23*, 2596–2608.

Grahofer, A., Letko, A., Häfliger, I. M., Jagannathan, V., Ducos, A., Richard, O., Peter, V., Nathues, H., & Drögemüller, C. (2019). Chromosomal imbalance in pigs showing a syndromic form of cleft palate. *BMC Genomics*, *20*, 349.

Hassold, T., & Hunt, P. (2001). To err (meiotically) is human: The genesis of human aneuploidy. *Nature Reviews Genetics*, *2*, 280–291.

Johnson, D. S., Gemelos, G., Baner, J., Ryan, A., Cinnioglu, C., Banjevic, M., Ross, R., Alper, M., Barrett, B., Frederick, J., Potter, D., Behr, B., & Rabinowitz, M. (2010). Preclinical validation of a microarray method for full molecular karyotyping of blastomeres in a 24-h protocol. *Human Reproduction*, *25*, 1066–1075.

Kern, A. D., & Schrider, D. R. (2018). *Diplos/hic: An updated approach to classifying selective sweeps*. G3. 2018;8:1959–1970.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks, Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. (pp. 1097–1105). Curran Associates Inc.

Martin, C. L., & Warburton, D. (2015). Detection of chromosomal aberrations in clinical practice: From karyotype to genome sequence. *Annual Review of Genomics and Human Genetics*, *16*, 309–326.

Martín Abadi, P. B., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Google Brain, Tensorflow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation, 2016, Savannah, GA, pp. 265–283.

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal snp and small-indel variant caller using deep neural networks. *Nature Biotechnology*, *36*, 983–987.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*, 60.

Sills, E. S., Li, X., Frederick, J. L., Khoury, C. D., & Potter, D. A. (2014). Determining parental origin of embryo aneuploidy: Analysis of genetic error observed in 305 embryos derived from anonymous donor oocyte ivf cycles. *Molecular Cytogenetics*, *7*, 68.

Silvestri, G., Canedo-Ribeiro, C., Serrano-Albal, M., Labrecque, R., Blondin, P., Larmer, S. G., Marras, G., Tutt, D. A. R., Handyside, A. H., Farré, M., Sinclair, K. D., & Griffin, D. K. (2021). Preimplantation genetic testing for aneuploidy improves live birth rates with in vitro produced bovine embryos: A blind retrospective study. *Cells*, *10*, 2284.

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*. Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 2015, arXiv:1409.1556.

Thorne, M. H., Collins, R. K., & Sheldon, B. L. (1991). Chromosome analysis of early embryonic mortality in layer and broiler chickens. *British poultry science*, *32*, 711–722.

Tšuiko, O., Catteeuw, M., Zamani Esteki, M., Destouni, A., Bogado Pascottini, O., Besenfelder, U., Havlicek, V., Smits, K., Kurg, A., Salumets, A., D'Hooghe, T., Voet, T., Van Soom, A., & Robert, V. J. (2017). Genome stability of bovine in vivo-conceived cleavage-stage embryos is higher compared to in vitro-produced embryos. *Human reproduction*, *32*, 2348–2357.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., Hakonarson, H., & Bucan, M. (2007). Penncnv: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome snp genotyping data. *Genome Research*, *17*, 1665–1674.

Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval*, *1*, 69–90.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bouwman, A C., Hulsegge, I., Hawken, R J., Henshall, J M., Veerkamp, R F., Schokker, D., & Kamphuis, C. (2023). Classifying aneuploidy in genotype intensity data using deep learning. *Journal of Animal Breeding and Genetics*, *140*, 304–315. https://doi.org/10.1111/jbg.12760