Comparaison d'algorithmes d'apprentissage automatique pour prédire le poids vif et les scores de pointage morphologique des veaux au sevrage à l'aide de l'imagerie 3D

DECHAUX T. (1), LEBRETON L. (1), DO Y. (1), EL JABRI M. (1), BRUYAS M. (2), DELATTRE L. (3), GILLE-PERRIER C. (4), ALLAIN C. (1)
(1) Institut de l'Elevage, 149 rue de Bercy, 75012 Paris
(2) Eliance, 149 rue de Bercy, 75012 Paris
(3) 3D Ouest, 5 rue de Broglie, 22300 Lannion
(4) Race de France, 149 rue de Bercy, 75012 Paris

RESUME
Lebreton *et al*. (2023) ont récemment démontré qu'il était possible de collecter des images tridimensionnelles de veaux au sevrage à l'aide d'un dispositif 3D adapté au phénotypage à haut débit et d'en extraire automatiquement des paramètres morphologiques (hauteurs, largeurs, volumes, surfaces...). Le but de cette nouvelle étude était de développer des modèles basés sur l'intelligence artificielle pour estimer le poids vif (PV) et les notes de pointage sur la base de paramètres morphologiques 3D préalablement estimés. Pour atteindre ces objectifs, 1114 veaux charolais âgés de 4 à 18 mois et pesant de 90 à 778 kg ont été scannés dans 14 élevages commerciaux. La plupart d'entre eux ont été scannés deux fois, générant 2 079 images 3D. Pour collecter des données de référence, chaque veau a été pesé sur une balance électronique (PV) et pointé visuellement par 3 techniciens expérimentés. La notation a conduit à l'estimation de 10 postes élémentaires qui ont ensuite permis le calcul de 2 notes synthétiques (notées de 1 à 100) : 1. Développement musculaire (MUS), relatif à la musculature globale de l'animal, et 2. Développement squelettique (SQE), relatif à sa structure corporelle. Pour prédire PV, MUS et SQE, divers algorithmes de machine learning (ML) ont été entraînés en utilisant 80 % des images et testés sur les 20 % restants. Les modèles ont été évalués à l'aide de l'erreur moyenne absolue (MAE) et de la corrélation de rang de Spearman (r_s). La répétabilité des prédictions a été évaluée par le r_s entre les estimations faites pour la première et la deuxième image. Pour le meilleur modèle, PV a été prédit avec un r_s de 0,97, un MAE de 12,1 kg et un MAPE de 4,2 %. La répétabilité r_s était de 0,98 entre les deux images. Pour le meilleur modèle ML, MUS et SQE ont été prédits avec des valeurs r_s de 0,78 et 0,75, respectivement, des MAE de 7,1 et 6,5 et des MAPE de 14,5 % et 11,9 %. Les r_s de répétabilité pour MUS et SQE étaient respectivement de 0,81 et 0,87. Les prédictions et la répétabilité obtenues pour MUS et SQE étaient supérieures à celles obtenues en moyenne par des pointeurs expérimentés lors des sessions annuelles de formation sur la race charolaise.

Comparison of Machine Learning Algorithms to predict body weight and linear scores of beef calves with 3D imaging

DECHAUX T. (1), LEBRETON L. (1), DO Y. (1), EL JABRI M. (1), BRUYAS M. (2), DELATTRE L. (3), GILLE-PERRIER C. (4), ALLAIN C. (1)
(1) Institut de l'Elevage, 149 rue de Bercy, 75012 Paris

SUMMARY
Lebreton *et al*. (2023) have recently demonstrated that it is possible to collect three-dimensional images of calves at weaning using a 3D device adapted for high-throughput phenotyping and to automatically extract morphological parameters (heights, widths, volumes, surfaces...). The goal of this new study was to develop models based on artificial intelligence to estimate body weight (PV) and linear scoring based on previously estimated morphological parameters. To achieve these goals, 1114 Charolais calves aged 4 to 18 months and weighing from 90 to 778 kg were scanned in 14 commercial farms. Most of them were scanned twice, resulting in 2079 3D images. To collect reference data, each calf was weighed on an electronic scale (PV) and scored visually by 3 experienced technicians. The scoring led to the estimation of 10 elementary scores which then allowed the calculation of 2 synthetic scores (rated from 1 to 100): 1. Muscle development (MUS), relating to the overall musculature of the animal, and 2. Skeletal development (SQE), relating to its frame. To predict PV, MUS, and SQE, various Machine Learning (ML) algorithms were trained using 80% of the images and tested on the remaining 20%. The models were evaluated using the Mean Absolute Error (MAE) and Spearman correlation (r_s). Repeatability of predictions was assessed by the r_s between the estimates made for the first and second image. For the best ML model, PV was predicted with a r_s of 0.97 a MAE of 12.1 kg and a MAPE of 4.2%. The repeatability r_s was 0.98 between the two images. For the best ML model, MUS and SQE were predicted with r_s values of 0.78 and 0.75, respectively, MAEs of 7.1 and 6.5 and MAPEs of 14.5% and 11.9%. The repeatability r_s for MUS and SQE were 0.81 and 0.87, respectively. The predictions and repeatability obtained for MUS and SQE were higher than those obtained on average by experienced scorers during annual accreditation sessions.

INTRODUCTION

Dans le secteur de l'élevage bovin en France, la sélection génétique repose principalement sur le phénotypage d'une large population animale (Griffon *et al*., 2017). Ce processus, indispensable pour les organisations de sélection (OS) implique un premier stade de collecte des phénotypes généralement autour du sevrage. Des techniciens du réseau Eliance ou des organisations de sélection effectuent cette collecte de données sur les exploitations, incluant la pesée des animaux et l'évaluation des caractères morphologiques par pointage visuel. Le pointage englobe 19 postes couvrant à la fois le développement musculaire et squelettique, ainsi que les aptitudes fonctionnelles. Le pointage est réalisé visuellement

par des techniciens formés pour chacune des races allaitantes françaises, comme décrit par Lajudie *et al*. (2014). Bien que le pointage ait joué et joue encore un rôle prépondérant dans le progrès génétique des bovins allaitants, il nécessite une formation considérable et est sujet à subjectivité importante de notations visuelles. Par conséquent, pour réduire les coûts de formation et d'agrémentation des opérateurs, et pour minimiser l'impact des biais humains sur les mesures, une automatisation du pointage des bovins viande au sevrage a été envisagée.

Afin de moderniser le processus, le projet PHENO3D a été initié, à partir d'une collaboration entre Eliance (la fédération française des entreprises de conseil et de service en élevage), Races de France (fédération française des organismes de sélection) et Idele (l'Institut de l'Élevage). L'objectif du projet est de tirer parti de la technologie d'imagerie 3D et de l'intelligence artificielle pour optimiser le processus de phénotypage en automatisant la mesure du poids et l'estimation des postes de pointage morphologique des veaux au sevrage (Bruyas *et al*., 2022). Un premier jalon de PHENO3D impliquait le développement d'un dispositif de numérisation 3D capable de capturer avec précision le profil tridimensionnel des veaux et d'en extraire des données morphologiques. La validation de cette technologie, utilisant une méthodologie similaire à celle décrite par Le Cozler *et al*. (2019a), a comparé des mesures réalisées manuellement sur des animaux vivants à celles dérivées des images 3D, montrant des résultats prometteurs (Lebreton *et al*., 2023). Cette validation réussie a permis d'ouvrir la voie aux phases de développement et d'application ultérieures.

1. MATERIEL ET METHODES

1.1. TECHNOLOGIE ET TRAITEMENT D'IMAGE

Le scanner 3D utilisé dans cette étude a déjà été décrit par Bruyas *et al*. (2022). Il s'agit d'un portique modulaire mesurant 3 x 2,5 x 0,7 mètres et incluant dix capteurs de profondeur (figure 1a). Ces capteurs synchronisent leurs processus d'acquisition de données pour générer des images 3D complètes de l'ensemble du corps des animaux scannés. Les animaux sont scannés en mouvement lorsqu'ils passent sous l'appareil en marchant ou en trottinant, permettant ainsi un phénotypage à haut débit. A partir de techniques de prétraitement et d'analyse de ces images 3D, des centaines d'indicateurs morphologiques sont ensuite automatiquement extraits (Do *et al*., 2024). Ces indicateurs comprennent par exemple la largeur de la hanche, la profondeur de la poitrine, la hauteur au garrot, la hauteur du sacrum, le volume corporel, la surface corporelle et beaucoup d'autres mesures (figure 1b). Tous ces indicateurs ont ensuite été utilisés pour construire les modèles de prédiction développés dans cette étude.

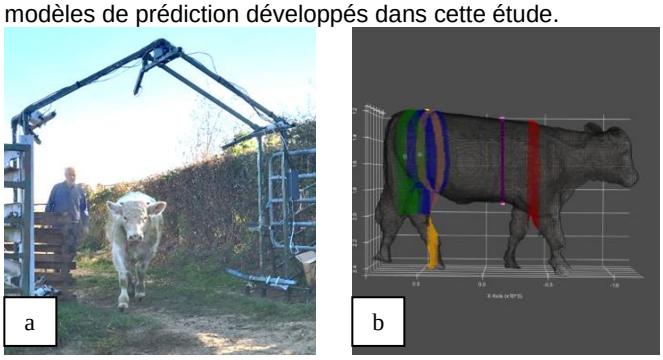

Figure 1 : Scanner 3D utilisé pour l'essai (a) et analyse d'image automatisée (b)

1.2. ANIMAUX ET DONNEES DE REFERENCE

Pour réaliser notre étude, 1114 veaux Charolais âgés de 4 à 18 mois et pesant de 90 à 778 kg ont été scannés dans 14 exploitations (tableau 1). La plupart d'entre eux ont été scannés deux fois, donnant un total de 2079 images 3D. Toutes ces images ont été stockées dans un cloud Microsoft Azure. Des mesures de référence ont été collectées sur ces mêmes animaux : chaque veau a été pesé sur une balance électronique pour mesurer son poids vif (PV) et pointé par 3 techniciens expérimentés. Le pointage a été effectué visuellement et a conduit à l'estimation de 10 postes élémentaires (notes allant de 1 à 10, score le plus bas = 1 et score le plus élevé = 10) caractérisant le développement musculaire et squelettique des animaux. Les postes évalués sont présentés sur la figure 2 ci-dessous.

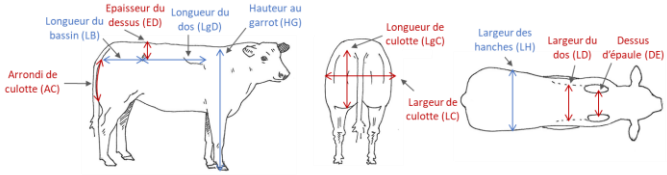

Figure 2 : Postes du développement squelettique (bleu) et musculaire (rouge) utilisés pour le pointage

Les 10 postes élémentaires ont ensuite été utilisés pour le calcul de 2 notes synthétiques (allant de 1 à 100) utilisées *in fine* pour la sélection génétique : 1. Développement musculaire (MUS), relatif à la musculature générale de l'animal, et 2. Développement squelettique (SQE), relatif à sa structure corporelle. Les formules de MUS (1) et SQE (2) sont présentées ci-dessous :

*(1) MUS = [(DE + LD + AC + LC + LgC*2 + ED)100]/70*
*(2) SQE = (LgD + LB + HW*2 + LH)*2*

Afin d'illustrer la variabilité morphologique des animaux de notre étude, la figure 3 ci-dessous montre des images 3D de veaux avec des morphologies extrêmes pour MUS et SQE. Les quatre images ci-dessous présentent des veaux d'environ le même âge mais avec des tailles (figure 3a) ou des musculatures (figure 3b) très différentes.

Figure 3 : Exemples d'images 3D de morphologies extrêmes pour des veaux âgés d'environ 6 mois. L'image (a) montre des veaux avec des notes allant de SQE=14 (gris clair) à SQE=94 (gris foncé). L'image (b) montre des veaux avec des notes allant de MUS=13 (gris clair) à MUS=94 (gris foncé).

Les détails concernant les statistiques descriptives des caractéristiques des animaux utilisés pour l'essai sont présentés dans le tableau 1.

n=1114	Âge	Poids	MUS	SKE
Moy.	221	294	56	59
ET	61	95	15	14
Min	44	90	10	12
Max	559	778	96	94

Tableau 1 statistiques descriptives des caractéristiques des animaux de l'étude

1.3. MODELES D'APPRENTISSAGE ET ANALYSE DES DONNEES

Dans notre base de données, un total de 2079 images 3D étaient initialement disponibles pour l'analyse. Cependant, certaines de ces images ont été supprimées en raison de divers facteurs affectant leur qualité, comme une mauvaise netteté, une position inappropriée de l'animal pendant la prise

d'image (comme des sauts ou des ruades) et des erreurs d'identification des animaux. Finalement, 1818 images ont été utilisées pour la prédiction du poids vif (PV) et 1575 images pour les prédictions des scores MUS et SQE.
Dans les deux cas, les modèles de prédiction ont été entraînés en utilisant 80 % des images et testés sur les 20 % restants, et aucune image du même animal n'a été incluse à la fois dans le jeu d'entraînement et le jeu de test.
Pour réaliser la prédiction du PV, 283 indicateurs ont été extraits des images 3D. Pour améliorer les performances prédictives, la méthode d'élimination récursive (RFE) a été utilisée. La RFE est une technique de sélection de variables utilisée en apprentissage automatique pour réduire la dimensionnalité des données en éliminant de manière itérative les caractéristiques les moins importantes du jeu de données. Cette technique a été appliquée sur le jeu de données d'entraînement en utilisant l'estimateur Random Forest et a abouti à la sélection de 61 indicateurs.
Après l'étape de sélection des indicateurs, quatre modèles d'apprentissage automatique (Extreme Gradient Boosting, Random Forest, SVM Linéaire et Régression Lasso) ont été entraînés sur les 1462 images du jeu de données d'apprentissage. Pour éliminer le surapprentissage, les modèles ont été entraînés en utilisant une méthode de validation croisée en 4 plis et avec 5 répétitions. Après cette phase d'entraînement, chaque modèle a été évalué sur le jeu de test de 356 images.
Pour prédire les notes synthétiques MUS et SQE, la même méthodologie a été utilisée. Tout d'abord, nous avons appliqué l'algorithme RFE pour sélectionner les indicateurs morphologiques les plus importants. Pour le score SQE, 16 indicateurs, y compris la prédiction du PV, ont été sélectionnés, tandis que pour MUS, 51 indicateurs ont été sélectionnés. Les mêmes 4 algorithmes ont été entraînés sur 1267 images avec une méthode de validation croisée en 4 plis, répétée 5 fois, et évalués sur le jeu de test de 308 images.
Les modèles ont été évalués en utilisant l'erreur absolue moyenne (MAE), l'erreur absolue moyenne en pourcentage (MAPE), l'erreur quadratique moyenne (RMSE) et la corrélation de rang de Spearman (r_s). La répétabilité des prédictions a été évaluée par la corrélation de Spearman entre les estimations faites pour la première et la deuxième image, lorsque des doublons étaient disponibles.

2. RESULTATS ET DISCUSSION

2.1. PREDICTION DU POIDS

Les performances des modèles sont présentées dans le tableau 2 ci-dessous. Parmi les modèles testés, l'Extreme Gradient Boosting s'est révélé être le plus performant. Le PV a été prédit avec une RMSE de 15 kg, une MAE de 11,3 kg et une MAPE de 4,0 % sur le jeu d'entraînement. La corrélation entre le poids réel et les prédictions du modèle ont atteint un r_s de 0,97 et un R^2 de 0,96.
La performance du modèle sur le jeu de test est similaire avec une RMSE de 15,6 kg, une MAE de 12,1 kg et une MAPE de 4,2 %. Pour le jeu de test, nous avons obtenu un r_s de 0,98 et un R^2 de 0,97. Le modèle n'a pas surappris du jeu d'entraînement, ce qui le rend plus fiable pour des applications réelles futures.
Le Random Forest (RF) a également montré des performances intéressantes, légèrement inférieures à celles de l'Extreme Gradient Boosting tandis que le SVM et la régression Lasso ont obtenus des performances inférieures.
La figure 3 montre la relation entre le PV prédit et le PV réel pour les jeux de données d'entraînement et de test. Avec un R^2 global de 0,96, les PV prédit et réels sont répartis autour de la droite y=x.

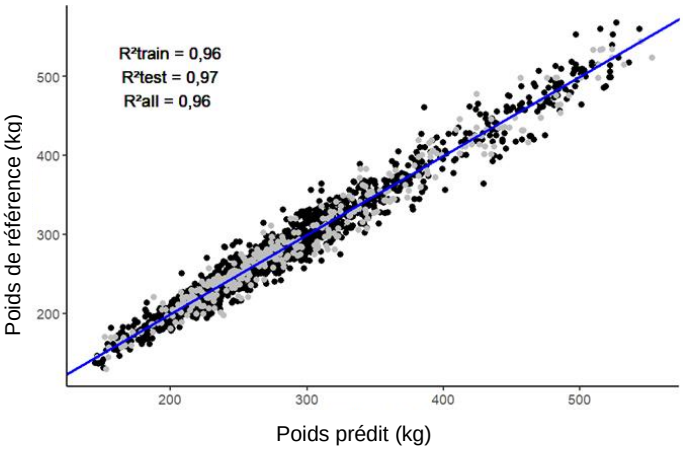

Figure 3 : PV prédit vs PV réel pour le jeu d'entraînement (noir) et le jeu de test (gris). La ligne bleue correspond à l'ajustement optimal.

La répétabilité du modèle a été évaluée pour assurer la cohérence des prédictions de poids lorsque différentes images du même individu. Elle a été calculée avec la corrélation de Spearman entre deux images du même bovin lorsque deux images 3D étaient disponibles. Nos résultats révèlent un niveau élevé de répétabilité du modèle avec un r_s de 0,98 pour les 738 bovins avec une image 3D dupliquée.

Nos résultats surpassent soit par les performances du modèle, soit par la puissance statistique liée au nombre d'animaux utilisés d'autres études récentes utilisant des techniques d'imagerie 3D et d'apprentissage automatique pour l'estimation du poids. Ruchay *et al.* (2022) ont obtenu un MAPE de 8,4 % dans la prédiction du PV pour des bovins de race Hereford. De même, Jang *et al*. (2020) ont rapporté un MAPE de 5,8 % dans leur étude sur la prédiction du PV, tandis que Grebreysus *et al*. (2023) ont obtenu un MAPE de 4,0 % pour l'estimation du PV des vaches Holstein et Jersiaises. Dans notre étude, le modèle le plus efficace a atteint un MAPE de 4,2 %, mais était entièrement automatisé et sur des animaux en mouvements. De plus, aucun de ces articles n'a exploré la répétabilité de leurs modèles de prédiction.

2.2. PREDICTION DES NOTES DE POINTAGE

Les performances des modèles sont présentées dans le tableau 2 ci-dessous. Le Random Forest et l'Extreme Gradient Boosting se sont révélés être les algorithmes les plus performants. Pour les prédictions des notes SQE, le Random Forest semble être plus précis, avec une MAE de 6,3, une MAPE de 11,3 % et une RMSE de 7,7 pour le jeu de test. La figure 4 représente la relation entre la note SQE de référence et la note SQE prédite par ce modèle. Le R^2 de 0,69 pour le jeu de test suggère une relation linéaire entre les prédictions et la référence. Des performances similaires ont été observées dans les jeux d'entraînement et de test, ce qui montre une bonne généralisation du modèle. Pour améliorer cette performance, il pourrait être bénéfique d'utiliser un certain nombre de notes plus extrêmes, notamment en dessous de 40, où le nombre d'animaux n'est pas suffisant dans notre étude. De plus, le modèle avait une bonne répétabilité avec un r_s de 0,91 entre les images dupliquées.
Pour les prédictions des notes MUS, l'Extreme Gradient Boosting était également le meilleur modèle avec une MAE de 7,1, une MAPE de 14,5 % et une RMSE de 9,1. La figure 5 montre une corrélation entre les prédictions et les références avec un R^2 de 0,65 pour le jeu de test. Le modèle avait également une bonne répétabilité avec un r_s de 0,81 entre les images dupliquées.
En France, chaque technicien du réseau national réalisant des pointages morphologiques est formé et testé chaque année par rapport à 3 techniciens de référence pour vérifier leur performance et leur éventuelle déviation.

Le coefficient de corrélation de Spearman moyen qu'ils atteignent en race charolaise est de 0,69 pour MUS et de 0,70 pour SQE. De plus, ils doivent démontrer leur répétabilité en atteignant un r_s de 0,75 pour MUS et de 0,78 pour SQE entre deux notations du même animal. Nos modèles ont significativement surpassé ces critères de performance, même si ce n'est pas sur le même jeu de données.

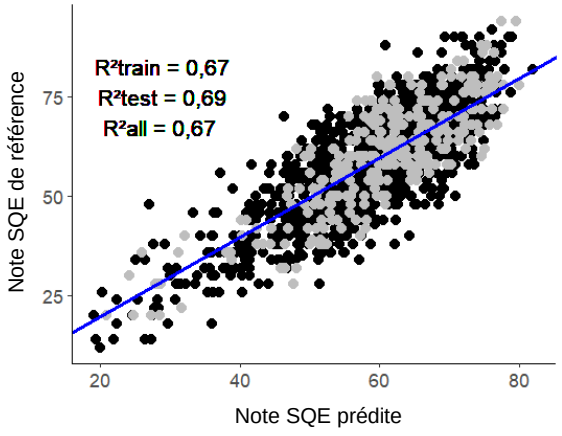

Figure 4 : Prédictions vs notes de référence SQE pour le jeu d'entraînement (noir) et le jeu de test (gris). La ligne bleue indique l'ajustement optimal

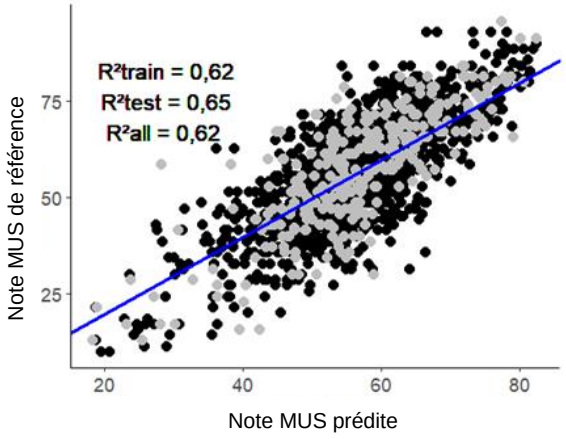

Figure 5 : Prédictions vs notes de référence MUS pour le jeu d'entraînement (noir) et le jeu de test (gris). La ligne bleue indique l'ajustement optimal

CONCLUSION

En conclusion, cette étude démontre la faisabilité de l'utilisation de l'imagerie tridimensionnelle couplée à des techniques d'intelligence artificielle pour une estimation précise du PV et des notes de pointage chez les veaux au sevrage. Grâce au développement de modèles d'apprentissage automatique, des prédictions robustes ont été réalisées pour le PV, le développement musculaire (MUS) et le développement squelettique (SQE), dépassant la performance moyenne des pointeurs du réseau national obtenue lors de leurs sessions de formation annuelles en race charolaise. La haute répétabilité des prédictions souligne la fiabilité de l'approche proposée, offrant une voie prometteuse pour une automatisation du phénotypage morphologique à terme.

Les auteurs expriment leur gratitude à tous les techniciens de notation et aux agriculteurs participants pour leurs contributions à l'essai. De plus, nous remercions le ministère français de l'Agriculture et APIS-GENE pour leur généreux soutien financier au projet PHENO3D.

Bruyas M., Lecomte C., Allain C., Lebreton A., Delattre L., Delouard J. M., Gautier J.M., Leudet O., Brochard M., Salvetti P., Griffon L., 2022. ICAR conf. 2024.
Do Y., Allain C., Dechaux T., El Jabri M., Delattre L., Bruyas M., Gillé-Perrier C., Lebreton A., 2024. ECPLF 2024.
Gebreyesus G., Milkevych V., Lassen J., Sahana G., 2023. Fron. Gen., 13, 947176.
Griffon L., Boulesteix P., Delpeuch A., Gion A., Guerrier J., Leudet O., Miller S., Saintilan R., Venot E., Tribout T., 2017. INRA Prod. Anim., 30(2), 107-124.
Hou Z., Huang L., Zhang Q., Miao Y., 2023. Comp. Elec. Agri., 213, 108184.
Jang D. H., Kim C., Ko Y. G., Kim Y. H., 2020. Jour. Bio. Eng., 45, 325-332.
Lajudie P., Berrechet P., Robergeot F., Deconchat P., Tournadre O., Roy M., Gout P., Roussie C., Hanriot J. C., Laloe D., 2014. https://idele.fr/fileadmin/medias/Documents/Manuel_pointage_vdef3.pdf
Lebreton A., Allain C., Delattre L., Loof J., Do Y., & Bruyas M., 2023. 74th EAAP Cong.
Le Cozler Y., Allain C., Caillot A., Delouard J. M., Delattre L., Luginbuhl T., Faverdin P., 2019a. Comp. Elec. Agri., 157, 447-453
Le Cozler Y., Allain C., Xavier C., Depuille L., Caillot A., Delouard J. M., and Faverdin P., 2019b. Comp. Elec. Agri., 165, 104977.
Ruchay A., Kober V., Dorofeev K., Kolpakov V., Gladkov A., Guo H., 2022. Agri., 12(11), 1794.

Tableau 2 : Performances des modèles pour la prédiction du PV et des notes de pointage synthétiques

Predicted traits	Jeux de données	Modèles Statistiques	*Random Forest*	*Extreme Gradient Boosting*	*Lasso Regression*	*SVM Linear*
PV	Jeux d'entraînement n=1462	MAE (kg)	11,7	11,3	12,9	12,8
		MAPE (%)	4,2%	4,0%	4,6%	4,6%
		RMSE (kg)	15,4	15,0	16,9	16,7
		r_s	0,97	0,97	0,96	0,96
		R^2	0,96	0,96	0,95	0,95
	Jeux de test n=356	MAE (kg)	12,4	12,1	13,6	13,3
		MAPE (%)	4,3%	4,2%	4,7%	4,5%
		RMSE (kg)	16,2	15,6	17,7	17,7
		r_s	0,97	0,98	0,97	0,97
		R^2	0,96	0,97	0,95	0,95
MUS	Jeux d'entraînement n=1267	MAE	7,6	7,1	7,6	7,6
		MAPE (%)	15,6%	14,5%	15,8%	15,5%
		RMSE	9,4	9,0	9,6	9,7
		r_s	0,73	0,75	0,72	0,73
		R^2	0,58	0,62	0,56	0,56
	Jeux de test n=308	MAE	7,4	7,1	8,0	7,9
		MAPE (%)	15,5%	14,5%	16,9%	16,7%
		RMSE	9,4	9,1	10,0	10,0
		r_s	0,77	0,78	0,74	0,74
		R^2	0,64	0,65	0,58	0,58
SQE	Jeux d'entraînement n=1267	MAE	6,3	6,1	7,2	7,2
		MAPE (%)	11,9%	11,3%	13,9%	14,0%
		RMSE	7,8	7,7	9,0	9,0
		r_s	0,78	0,79	0,74	0,73
		R^2	0,67	0,68	0,56	0,55
	Jeux de test n=308	MAE	6,3	6,5	6,9	7,0
		MAPE (%)	11,3%	11,9%	12,9%	13,1%
		RMSE	7,7	8,0	8,5	8,5
		r_s	0,78	0,75	0,76	0,75
		R^2	0,69	0,67	0,62	0,62