



HAL
open science

Étude du comportement maternel des truies

Léo Denis

► **To cite this version:**

| Léo Denis. Étude du comportement maternel des truies. Informatique [cs]. 2020. hal-02957989

HAL Id: hal-02957989

<https://hal.inrae.fr/hal-02957989>

Submitted on 5 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude du comportement maternel des truies Rapport de stage

Léo DENIS

1. Introduction

1.1. Présentation

J'ai effectué mon stage entre le 17 février et le 23 juin 2020 au centre de l'INRAE à Castanet-Tolosan, dans le laboratoire GenPhySE. L'INRAE (Institut Nationale de Recherche pour l'Agriculture, l'Alimentation et l'Environnement) est le premier organisme de recherche spécialisé dans ces domaines. Le laboratoire GenPhySE (Génétique, Physionomie et Système d'Élevage) se spécialise dans la génétique des animaux et les effets de l'environnement sur les phénotypes. J'ai effectué mon stage sous la responsabilité de Laurianne Canario, Mathieu Bonneau et Yann Labrune dans le cadre de leur travail sur l'activité posturale des truies d'élevage. A partir du 16 mars, dû à la pandémie de Covid-19, les locaux ont fermé et mon stage s'est poursuivi en télétravail jusqu'à sa conclusion.

1.2. Contexte et Problématique

La problématique du stage est de mettre au point des outils de suivi automatisé du comportement maternel de truies en conditions expérimentales, c'est à dire bloquées entre des barrières, qui n'ont donc pas de liberté de mouvement. Aujourd'hui, il y a une demande croissante de la société pour que les animaux soient capables d'exprimer leurs comportements naturels et cela nécessite de changer de système d'hébergement pour leur apporter plus de liberté. Dans le cas de la production porcine, plus de 90% des truies sont encore totalement bloquées pendant la lactation. Mon stage vise à caractériser le comportement postural des truies dans cet environnement témoin. La problématique du stage est de mettre au point une méthode de suivi automatisée du comportement maternel de truies tandis qu'elles sont bloquées entre des barrières, et n'ont donc que la liberté de changer de position.

Initialement, ce type de logement a été développé pour empêcher que la truie n'écrase des porcelets. Cependant, certaines truies ont une activité posturale qui les amène à écraser des porcelets, que ce soit volontairement (réaction de néophobie), ou involontairement (mouvements trop brusques, par exemple dus à des problèmes aux pattes). C'est ce problème de mortalité des porcelets qui est le moteur premier de ce travail, car elle est bien plus élevée chez les



FIGURE 1: Capteur embarqué sur la truie

truies en liberté que chez ceux confinés (c. 2x plus) d'après les recherches de l'INRAE et certains articles Pedersen (2015). D'autres sources, en revanche, ne constatent aucune différence significative (Lambertz et al., 2015) (Singh et al., 2017). Avec l'évolution des systèmes, il est donc nécessaire de repérer les truies ayant un comportement maternel favorable à la survie du plus grand nombre qui pourront donc se voir accorder plus de liberté.

Nous disposons d'enregistrements vidéo réalisés sur des truies sur une période de 24 à 48h. L'objectif sera de mettre au point et comparer différentes méthodes pour analyser les images extraites à haute fréquence et définir le patron d'activité.

1.3. Animaux et données zootechniques

Les truies étudiées sont issues d'une population Large White sur Unité expérimentale du Magneraud GENESI (Charentes-Maritimes). Les truies entrent en maternité environ 1 semaine avant la mise bas, et ensuite pendant la lactation, sont suivies avec des enregistrements vidéo sur une période de 24h à 96h. La lactation dure 4 semaines, les enregistrements débutent à X jours après la mise bas. Les truies sont bloquées de leur entrée en maternité à fin de la lactation Deux mesures d'intérêt sont prises pour la mise au point : poids des truies juste avant l'entrée en maternité, hauteur au garot mesurée au moment de l'enregistrement

1.4. État de la recherche

● = CD ; ● = CG ; ● = CV ; ● = A ; ● = D

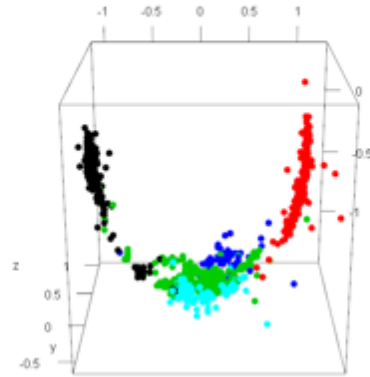


FIGURE 2: Visualisation des données accéléromètre. Les positions indiquées sont dans l'ordre : couché droit, couché gauche, couché ventre, assis et debout.

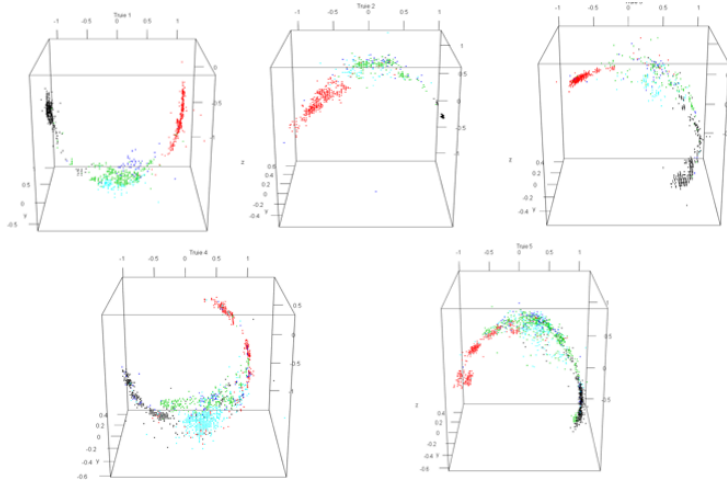


FIGURE 3: Comparaison des résultats des capteurs pour cinq truies différentes

Des tests avaient été effectués avec trois types d'accéléromètre (dont un Smartphone) afin d'observer et de comparer les résultats. Le premier a été choisi car il présentait un meilleur compromis entre les capacités de détection et l'autonomie du système.

Les données apparaissent comme sur la Figure 2

Cette méthode permet déjà de bien isoler les positions, en particulier les positions couché droit et couché gauche. Cependant, la variabilité dans la façon dont le harnais contenant le capteur est attaché mène à une trop grande variabilité dans les résultats. La figure 3 montre que l'orientation des données diffère complètement d'une truie à l'autre.

2. Traitement vidéo

Dans le cadre de ce stage, je n'ai travaillé que sur le traitement vidéo. A cause de problèmes techniques, les nouvelles données accéléromètre n'étaient pas disponibles. La première étape était de travailler sur la méthode de segmentation pour isoler la truie du reste des éléments de l'image afin d'obtenir un unique "cluster" représentant l'animal. Le résultat est donc une image binaire, où le cluster, c'est à dire la truie, apparaît en blanc, et le reste (décors et porcelets) apparaît en noir. L'idée est ensuite de calculer des variables permettant de caractériser la forme du cluster pour déterminer la position de la truie. Voici donc les étapes de mon travail sur le traitement vidéos.

2.1. *Region of Interest*

Les vidéos sont filmées avec une caméra BASCOM fixe placée au dessus de la truie. Celle-ci est entravée par des barres de métal placées dans une case dans laquelle peuvent circuler les porcelets. L'éclairage et l'angle de caméra peuvent varier, d'une journée ou d'une truie à l'autre.

Pour la détection de la truie, une grande portion de l'image, en dehors de barrières, est inutile puisque la truie ne sort pas de ses restrictions. La première étape fut donc de rogner l'image pour définir une zone d'intérêt (region of interest, par la suite abrégé en ROI). La truie se meut de haut en bas selon l'axe vertical de l'image (notamment pour se nourrir) et peut occuper toute la hauteur de l'image. Il suffit donc de rogner sur la largeur. En premier lieu j'ai centré cette ROI au centre de la largeur de l'image. L'enclos, et donc la truie, n'étant pas forcément au centre du cadre de la caméra, j'ai par la suite utilisé le centre de la largeur du cluster principal calculé sur la première frame de la vidéo.

Cette méthode ne présente toujours pas des résultats optimaux, la variabilité de la position initiale de la truie fait qu'un morceaux de l'enclos est souvent oublié. Il fut donc finalement décidé de noter manuellement la zone d'intérêt pour chaque vidéo. Les limites de l'enclos selon l'axe horizontale sont notées sur un document texte qui est lu par le programme au début de l'analyse.



FIGURE 4: Truie des premières vidéos à gauche. Truie dans l'étude du stage à droite.

2.2. Seuillage

Il convint ensuite de segmenter la truie. J'ai débuté par un filtrage sur la couleur. J'ai donc converti l'image en HSV, sélectionné un intervalle à conserver dans la teinte (dans les rouges car la truie est rose), et j'ai seuillé ce qui était dans cet intervalle. J'ai développé cette méthode avant le début officiel du stage sur des vidéos issues d'un jeu de données différent. Pour une première méthode rudimentaire, cela fonctionnait extrêmement bien. On distinguait clairement la forme de la truie.

Malheureusement, quand j'ai eu accès aux vidéos de l'étude concernée, cette méthode ne fonctionnait plus du tout. C'est certainement dû à une différence de luminosité ou dans les couleurs de la caméra (on voit bien sur l'image que les couleurs de la truie ressortent moins, elle paraît plus blanche)

Pour sa simplicité en termes de calcul j'ai mis au point une technique de segmentation basée sur du seuillage. J'ai testé différentes lois de seuillage dans différents espaces de couleurs (HSV et RGB), mais finalement, un simple seuillage sur les niveaux de gris s'est avéré la méthode la plus pertinente. J'ai donc commencé à travailler sur une image binaire issue de ce seuillage avec un seuil choisi manuellement. Plus tard, j'ai utilisé la méthode d'Otsu pour estimer le seuil. (Liu and Yu, 2009) Méthodes fournies par open-cv :

```
cv2.threshold(cv2.cvtColor(roi, cv2.COLOR_BGR2GRAY), 0, 255, cv2.THRESH_BINARY+cv2.THRESH_OTSU)
```

Cette image seuillée est ensuite soumise à une fonction d'érosion afin d'éliminer en partie le bruit, puis de dilatation pour restaurer la forme de la truie.

2.3. Cluster principal

Cette méthode de seuillage permet donc de visuellement discerner la truie, mais garde aussi beaucoup d'éléments non-pertinents. L'image est composée de "taches", ou clusters, blanches (différents éléments du décors, les barrières, les porcelets, du bruit...). Il s'agit maintenant de déterminer le cluster correspondant à la truie et les clusters correspondant aux éléments du décors. Pour ceci, nous avons émit l'hypothèse que le cluster correspondant à la truie est celui présentant l'aire la plus importante. J'ai donc travaillé sur une méthode pour isoler ce cluster et supprimer tout le reste. Pour ce faire, deux méthodes furent testées.

La première utilise les fonctions de détection de contours d'opencv. Celle-ci renvoie une liste de contours présents sur l'image. Elle remplit ensuite ce contour pour en calculer l'aire. Le contour ayant l'aire la plus grande est conservé.

La seconde utilise la méthode de Connected Components, fournie par la fonction labels de scipy.ndimage. Celle ci renvoie tous les blocs de pixels connectés. Il suffit alors de sélectionner celui présentant le plus d'éléments.

Ces deux méthodes donnent le même résultat mais sont malheureusement assez lourdes. J'ai fait un test de rapidité sur les deux méthodes, en moyennant sur les même 1000 frames :

Méthode 1 : 1.1111 secondes par frame.

Méthode 2 : 1.1778 secondes par frame.

A ma surprise, la première méthode que j'avais mise au point était légèrement plus rapide. Par la suite , on a redimensionné l'image. En effet, les images reçues avaient une résolution de 720*1280 pixels, soit beaucoup de données qui ralentissaient considérablement les calculs sans apporter d'information pertinente en plus. Après avoir divisé la taille de l'image par deux, les même tests donnaient ces résultats :

Méthode 1 : 0.4931 secondes par frame.

Méthode 2 : 0.6229 secondes par frame.

L'écart se creuse, la méthode 1 est donc bien à privilégier.

Malheureusement, cette méthode n'est pas encore idéale. Tant qu'ils sont connectés par des pixels, d'autres éléments de l'image continuent d'être inclus dans le Cluster. C'est vrai notamment pour les porcelets problème mineur car ils sont petits et n'altèrent pas la forme générale du cluster s'ils ne sont pas présents en masse. S'ils le sont, cela traduit de manière générale une période d'allaitement, et ils sont donc un facteur de décisions positif pour prédire ces moments. Mais c'est surtout un problème pour les barrières. De couleur claire, elles peuvent se mêler au cluster et, par connexions interposées, faire apparaître beaucoup d'éléments non-désirables dans l'image. Pour réduire ces effets, une méthode testée est de générer une ellipse autour de la truie, à partir du centre du cluster, de sa largeur et de sa hauteur. Les pixels extérieurs à cette ellipse



FIGURE 5: les différentes étapes de la segmentation. 1 : La frame originale. 2. Seuillage binaire. 3.Cluster principal. 4. Ellipse générée. 5. Cluster principal dans l'ellipse.

sont alors mis à 0, réduisant ainsi les éléments parasites. Les différentes étapes de cette méthode sont illustrées sur la figure 5

2.4. Extraction des variables

Une fois obtenue une segmentation satisfaisante de la truie, il faut en extraire les informations nécessaires à l'analyse statistique.

J'utilise la fonction de scikit image `simage.mesures.regionprops` pour extraire ces informations. Celle-ci prend une image "labellée" (ici, un label unique, le cluster étudié) et renvoie un certain nombre d'informations sur celle-ci. J'extraits les 11 variables suivantes : area, convex area, eccentricity, equivalent diameter, extent, major axis length, minor axis length, orientation, euler number, centroid x et centroid y.

Ces variables sont extraites sur le cluster entier, mais pas uniquement. La supposition était qu'étudier la valeur de ces variables sur des parties de l'image apporterait d'autres informations. J'ai ainsi découpé le Cluster en trois horizontalement, pour isoler la tête, le corps et l'arrière de la truie. Étudier la forme et l'orientation de la tête en particulier peut par exemple indiquer plus précisément si la truie est debout ou allongée. Accorder plus d'importance à l'arrière de la truie permet par exemple de déterminer si elle est assise ou allongée. Le découpage considéré est le suivant : 1/5ème du cluster en partant du haut est a tête. De 1/5ème à 3/5ème de la hauteur du cluster est le corps et les deux dernières cinquièmes sont l'arrière. On découpe également le cluster en deux verticalement selon le milieu de la ROI. Cette information pourrait ainsi permettre de différencier sur quel flanc la truie est allongée. Toutes les variables sont donc extraites sur chacune de ces images, soit 66 variables

En addition de ces variables, d'autres valeurs sont évaluées sur le cluster entier. On mesure ainsi la proportion de pixels blancs de la tache à trois niveaux horizontalement (correspondant à la largeur des épaules [aux 3/10 de la hauteur

du cluster en partant du haut], des flancs [à la moitié] et de l'arrière [aux 3/4]), sur trois niveaux verticalement et la proportion de pixels blancs sur les deux diagonales de la ROI. Enfin, le périmètre du cluster principal est évalué, pour un total de 75 variables.

Pour comparer de manière plus pertinentes des truies différentes, le poids de la truie à la pesée avant l'étude est ajouté. C'est donc une variable qui est constante pour une truie donnée. Cette donnée n'est pas déterminée pendant l'analyse vidéo, il faut donc noter à la main le poids de la truie sur un document texte qui sera lu par le programme.

2.5. Organisation des vidéos

Une fois définies les fonctions de traitement vidéo. Il a fallu organiser le code de manière à automatiser la lecture des vidéos, leur traitement et l'enregistrement des variables.

Les truies sont filmées journée par journée. Une journée d'observation est enregistrée selon une nomenclature suivant cet exemple :

Bande 2001\Serie3\25-01-20 truie2-810596

Les deux premiers dossiers sont organisés selon les différentes séries d'études organisées. Il y a actuellement trois "bandes" (2001, 2002, 2003) contenant chacune 2 à 6 "séries" (contenant elles-même un nombre variable de truies). Le troisième dossier contient l'enregistrement d'une truie sur une journée. Le nom du dossier indique en premier la date de l'enregistrement, puis le numéro de la truie dans cette étude (ici 2), et enfin la désignation de la truie dans l'élevage (nombre à 6 chiffres, ici 810596) Dans ce dossier se trouve l'enregistrement de la truie, découpé en vidéos de durée variable mais environ une heure chacune. Les vidéos sont intitulées suivant cet exemple : 0 8 IPC2 20200125064950

'0 8 IPC' est une norme commune à toutes les vidéos. Le '2' est le numéro de la truie dans la série. Le reste est la date et l'heure du début de la vidéo (ici le 25 janvier 2020 à 06h49 et 50 secondes)

Lorsque la vidéo est analysée, des documents textes sont générés et enregistrés dans un dossier "DATA" créé dans le dossier de la truie.

Les vidéos étant filmées sur toute la journée, on a donc des vidéos de la truie de jour et de nuit. Les vidéos de nuit son filmées en infrarouge et demandent donc un traitement différent des vidéos de jour. Mon stage s'est concentré sur les vidéos de jour, car les truies y sont plus actives. J'isole et ignore donc toutes les vidéos de nuit avant de procéder à l'analyse vidéo.

En nombre limitées au début, j'ai eu accès à de nouvelles vidéos annotées au fur et à mesure du stage. A la fin du stage, je disposais de données sur 13 truies. Dix d'entre elles sur une journée, Deux sur deux journées et une sur trois journées.

2.6. Analyse visuelle

Afin de faire des analyses statistiques sur les données, une dernière chose indispensable était d'avoir la vérité terrain sur la position des truies analysées. Trois observateurs ont analysées les enregistrements complets utilisés dans cette

mise au point de vue pour définir les positions suivantes, définies par l'éthogramme suivant, accompagné d'un descriptif visuel des positions (Annexe x).

Les documents donnant la position de la truie au fil de la journée se présentent sous la forme de fichiers classeurs à deux colonnes. La première colonne donne la frame du changement de position. Les positions possibles étaient initialement :

- 0 : couché droite
- 1 : droite allaitement
- 2 : couché gauche
- 3 : gauche allaitement
- 4 : ventre
- 5 : assis
- 6 : debout
- 7 : genoux

Les positions couchées gauche et droite s'appliquent quand la truie est couchée avec les mamelles cachées sous elle. Les positions allaitement désignent les truies couchées sur le flanc avec les mamelles exposées, rendant l'allaitement possible.

Ces documents sont enregistrés de manière à coïncider avec le découpage en vidéos d'une journée d'enregistrement de la truie. La manière donc ils sont créés présente un défaut : étant donné que seuls les changements de position sont enregistrés, la position de la truie au début de chaque vidéo n'est pas donnée par le document. Pour y remédier, la fonction qui lit ces classeurs prend en argument la position de départ, et retourne la position finale. Ainsi, lorsque les vidéos d'une journée sont traitées à la suite et dans l'ordre, la position de départ de la truie est transmise (puisque une truie démarre forcément une vidéo dans la position où elle a terminé la précédente)

3. Analyse statistique

3.1. Traitement des variables

On a initialement extrait un maximum de variables pour récupérer autant d'information que possible. Ensuite, nous avons sélectionné celles qui apportaient une information pertinente et retiré certaines qui étaient redondantes ou non pertinentes. Dans un premier temps, il s'agit de déterminer les variables aberrantes, celles qui n'apportent aucune information cohérente. La démarche a consisté pour chaque truie étudiée et pour chaque position, à calculer la moyenne de chaque variable. Puis, pour une position donnée, on calcule la variance de ces moyennes entre truies. Si cette variance dépasse un certain seuil, on considère la variable comme trop erratique et elle est supprimée. De la même manière on calcule d'abord les variances, puis on fait la moyenne des variances pour chaque position. Si en moyenne, la variance d'une variable dépasse un certain seuil, c'est que la variable a un comportement aberrant par rapport au changement de position. Fig. (6). Fig. (7)

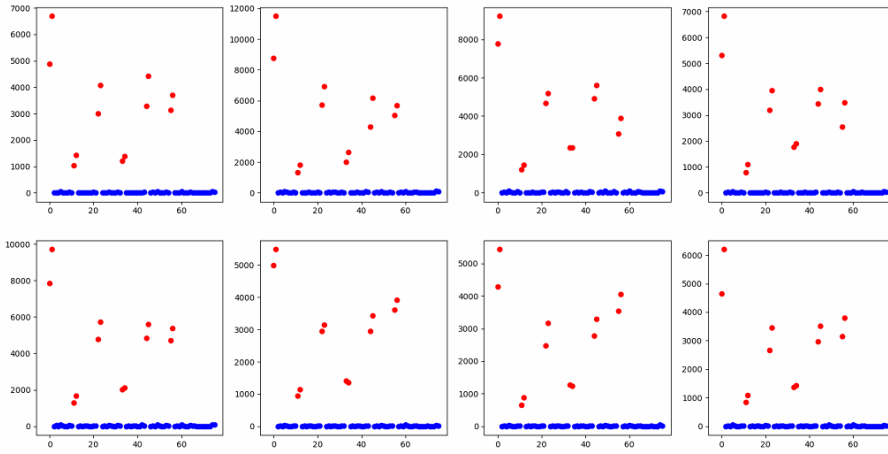


FIGURE 6: Variance des moyennes de l'erreur pour chaque truie en fonction de la position. Les variables dont la moyenne dépasse le seuil sont affichées en rouge, les autres en bleu.

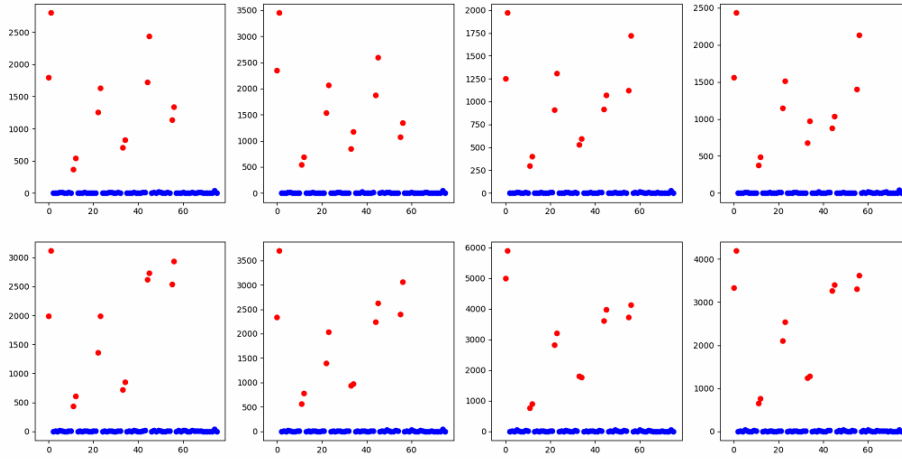


FIGURE 7: Moyenne des variances de l'erreur pour chaque truie en fonction de la position. Les variables dont la variance dépasse le seuil sont affichées en rouge, les autres en bleu.

Méthode	Nb. estimateurs	Erreur 1	Erreur 2
Random	10	0.0213	0.3126
Forest	100	0.0186	0.2973
Extra Trees	10	0.0189	0.3113
Classifier	100	0.0169	0.2883
Ada Boost	10	0.490	0.849
Classifier	100	0.274	0.633
Gradient	10	0.195	0.580
Boost	100	0.055	
Linear Discriminant Analysis		0.231	0.601
Quadratic Discriminant Analysis		0.130	0.434
Naive Bayes		0.533	0.926

TABLE 1: Erreur de chaque méthode de classification pour les deux méthodes de sélection (erreur 1 pour la méthode 1, erreur 2 pour la méthode 2) avec 30% d'apprentissage

Le seuil choisi est 5% de la valeur maximum. On voit sur les graphiques que les variables présentant des valeurs aberrantes le fait de manière nette, on les supprime donc.

En procédant ainsi, j'avais peur de supprimer trop de variables. En effet, certaines variables peuvent être pertinentes pour certaines positions mais pas pour d'autres. Mais en testant la méthode à différentes étapes du stage, supprimer ces variables ne présentait jamais de changement significatif dans l'erreur produite par la méthode. Elle n'avait donc ni d'effet positif, ni négatif sur les résultats, mais les supprimer allège les calculs.

3.2. Séparation des données

Afin de préparer l'utilisation de méthodes de classifications, il a fallu déterminer comment séparer les données en jeu d'apprentissage et en jeu de test. Du à la faible quantité de vidéos annotées disponibles au début du stage, les premières approches furent d'observer les performances de la prédiction avec les ensembles d'apprentissage et de test issus de la même vidéo.

Deux approches furent envisagées. La première consiste à prendre les données du jeu d'apprentissage au hasard en respectant les proportions de chaque position. La seconde consiste à prendre une plage de données continue, suffisamment longue pour probablement contenir toutes les positions (Le matin est privilégié car la truie y est plus active et présente donc plus de changements de position).

3.3. Méthodes de classification

Le parti pris en début de stage était de se focaliser sur des méthodes d'apprentissage et de classification simples, sans réseaux de neurones. Il fallait donc sélectionner la méthode de classification à appliquer. Plusieurs méthodes furent ainsi envisagées et testées avec les deux méthodes. Les résultats sont présentés dans le tableau 1

Pour les deux découpages, les méthodes utilisant et comparant des arbres de décision, Random Forest et Extra Trees, sont celles qui donnent les meilleurs résultats. Cela concorde avec les résultats des études précédentes.

Ces résultats furent obtenus tôt dans le stage, avec une méthode de segmentation encore incomplète et des données limitées. Des tests faits au fur et à mesure de la progression de l'algorithme confirment ce résultat. Dans la suite du stage, les tests sont fait avec une méthode de Random Forest.

On remarque surtout que malgré que l'algorithme ne soit pas abouti, la première méthode de découpage donne d'excellents résultats. En réalité, ces résultats sont faussés et rejoignent un problème déjà rencontré lors des premières études. Ces résultats sont dus à un surapprentissage des données. En effet, les données sont fortement corrélées dans le temps. En prenant les valeurs d'entraînements aléatoirement dans toute la vidéo, on s'assure de tester sur des images quasiment identiques au jeu d'entraînement, d'où les excellents résultats. On constatera plus tard que ce modèle ne fonctionne pas quand appliqué à une autre truie.

3.4. *Filtre median*

Par la suite, on appliquera également un traitement supplémentaire aux données prédites. En effet, si la fréquence des frames est élevée, la nature même des données suggère qu'il n'y a pas de changement ponctuels de la variables à prédire. C'est à dire que par exemple, si une truie est allongée, elle ne va pas être debout à la frame suivante puis allongée à celle d'après. Après la prédiction, j'applique donc un filtre median au vecteur de données prédites (ordonnées).

Les tests faits en ce sens au cours du stage me donnaient des résultats qui s'améliorait faiblement mais constamment avec l'augmentation de la taille du filtre, jusqu'à une taille d'environ 500. Cependant, des tests plus rigoureux faits en fin de stage donnent des résultats contradictoires, montrant que l'application du filtre a plutôt tendance à augmenter l'erreur. Voici un exemple d'effet du filtre sur l'erreur, appliqué à la prédiction sur les données de la truie sur trois journées, avec 30% de données d'entraînement prises en début et en fin de journée.

Taille du filtre : Erreur de la méthode.

0 : 0.5626243941274264
5 : 0.5627531786358209
11 : 0.5629170861919591
21 : 0.5624370712061254
55 : 0.5644273772449482
81 : 0.5635610087339312
151 : 0.5644625002926921
301 : 0.5652000842953145
501 : 0.5696021729458871

Le filtre médian est donc abandonné par la suite.

Truie n.	1	% 1	2	% 2	3	% 3
Erreur globale	0.585		0.664		0.690	
Erreur couchée droite	1	0.1	0.999	5.38	0.934	6.2
Erreur allaitement droite	0.408	19.4	0.957	3.5	1	0.2
Erreur couchée gauche	0.747	15.5	0.879	4.7	0.150	10.6
Erreur allaitement gauche	0.879	3.7	0.975	5.7	0.983	38.4
Erreur couchée ventre	0.999	11.1	0.610	21.9	0.030	18.6
Erreur assise	1	0.4	0.726	6.9	0.131	8.0
Erreur debout	0.015	48.4	0.002	51.1	0.022	17.9
Erreur genoux	0.998	1.35	0.944	0.9	0.784	0.2

TABLE 2: Erreur de chaque position pour les trois truies testées avec 30% des données en apprentissage, suivi du pourcentage de temps passé dans chaque position dans le jeu d'apprentissage.

3.5. Gestion des positions

Le patron d'activité donné initialement présentait 8 positions. La question se posait alors de savoir si chacune de ces positions étaient pertinente ou si certaines nuisaient à la prédiction. Pour observer la qualité de la prédiction pour chaque position, j'ai effectué trois tests de classifications distincts sur les données des trois truies pour lesquelles on a plus d'une journée d'enregistrement. Ces tests ont été faits avec 30% des données, prises en début et fin de journée, comme jeu d'apprentissage, et le reste en test. Les résultats sont reportés dans le tableau 2. La figure 8 donne les matrices de confusion des trois tests.

En étudiant le tableau, la première observation est l'importance de la quantité de données disponibles dans une position. Dans la plupart des cas, plus le temps passé dans un position est élevé, plus l'erreur est faible. Ce n'est cependant pas le seul facteur, certaines variables présentant même le comportement inverse (genoux, allaitement gauche). Par ailleurs la quantité de données elle-même n'est pas garante de meilleurs résultats. La truie 3 est celle étudiée sur trois jours (contre deux pour les deux autres) et est celle qui présente la plus grosse erreur. Ces informations suggèrent déjà que l'analyse présente de grandes différences d'une truie à l'autre.

Si on observe que la position debout est de loin la mieux prédite, l'étude des matrices de confusion nous montre que c'est aux dépends des autres prédictions. Pour les deux premières truies en effet, quasiment toutes les positions sont fortement mal prédites comme "debout".

Les trois positions "couchées" (sans allaitement) sont toutes mal prédites. Après discussion sur la réalité de ces positions, il s'est avéré que non seulement elles étaient visuellement très proches, mais que les distinguer ne présentait pas une importance capitale, contrairement aux positions exposant les mamelles. Il fut donc décider de fusionner ces trois positions en une seule, qu'on nommera désormais "couchée". Le même test est réalisé après cette fusion de position, reporté dans le tableau 3 et la figure 9.

	0	1	2	3	4	5	6	7
0	0	0.375	0	0	0	0	0.625	0
1	0	0.591598	0	0	0.000803729	0	0.407062	0.00053582
2	0	6.43418e-05	0.252477	0.00386051	0	0	0.739094	0.00450392
3	0	0	0.0882276	0.120779	3.28472e-05	0	0.790205	0.000755485
4	0	0.0209462	0.0730921	0.00117182	0.00051267	0	0.904277	0
5	0	0	0	0	0.0421394	0	0.957861	0
6	0	0.00526356	0	0.00560934	0.00376518	0.000153681	0.985131	7.68403e-05
7	0	0.518272	0.00332226	0	0.00664452	0	0.4701	0.00166113
	0	1	2	3	4	5	6	7
0	0.00028023	0.037364	0	0.455	0.0561394	0	0.12956	0.321657
1	0	0.0425017	0	0.315169	0.042111	7.81281e-05	0.257705	0.342435
2	0	0	0.121115	0.00214362	0.000214362	0	0.875456	0.00107181
3	0	0.000237567	0.00886918	0.0250238	0.018847	0.00118784	0.944092	0.00174216
4	0.119277	0.0366968	0.0020762	0	0.38965	0.0323368	0.292432	0.12753
5	0.0014301	0.000893815	0.00196639	0	0.00446907	0.273865	0.717197	0.000178763
6	0	0	0	0.000278193	0.00104322	6.95483e-05	0.997879	0.000730257
7	0	0	0.0846561	0.0026455	0.042328	0	0.814815	0.0555556
	0	1	2	3	4	5	6	7
0	0.0655721	0.000103263	0.000206526	0.00216853	0.921107	0.00351095	0.00733168	0
1	0	0	0.0123245	0.812301	0.0293156	0.00598277	0.139917	0.000159541
2	0	0	0.849454	0	0.15033	0	0.000216146	0
3	0	0	0.977936	0.0171126	0.000434329	0.000477762	0.00403926	0
4	0.00635401	0	0.00726172	0.000453858	0.969894	0.014826	0.00121029	0
5	0	0	0.00282965	0	0.0622524	0.86927	0.065648	0
6	0	0	0	0	0.000305686	0.0212961	0.978398	0
7	0	0	0.0263158	0	0.0894737	0	0.668421	0.215789

FIGURE 8: Matrices de confusion des trois tests avec 30% des données en apprentissage.

Truie n.	1	% 1	2	% 2	3	% 3
Erreur globale	0.690		0.609		0.516	
Erreur couchée	0.928	26.7	0.682	32.0	0.007	35.4
Erreur allaitement droite	0.628	19.4	0.956	3.5	1	0.2
Erreur allaitement gauche	0.840	3.7	0.981	5.7	0.798	38.4
Erreur assise	1	0.4	0.689	6.9	0.145	8.0
Erreur debout	0.074	48.4	0.002	51.1	0.009	17.9
Erreur genoux	0.995	1.35	0.952	0.9	0.821	0.2

TABLE 3: Erreur de chaque position pour les trois truies testées avec 30% des données en apprentissage, suivi du pourcentage de temps passé dans chaque position dans le jeu d'apprentissage, après fusion des positions couchées.

	0	1	2	3	4	5
0	0.0721713	0.0180599	0.0150499	0	0.894685	3.42044e-05
1	0.00289343	0.372287	0	0.000107164	0.624123	0.000589401
2	0.0546249	0	0.159802	0	0.785245	0.000328472
3	0.128579	0.0983252	0	0	0.773096	0
4	0.0426848	0.028623	0.00165207	0.000614723	0.926118	0.000307361
5	0.00996678	0.461794	0.0116279	0	0.511628	0.00498339
	0	1	2	3	4	5
0	0.317939	0.0544749	0.3362	0.00348463	0.170526	0.117375
1	0.0130083	0.043908	0.583226	0	0.0209774	0.33888
2	0.602154	7.91891e-05	0.0189262	0.000633513	0.376148	0.00205892
3	0.0516625	0.000178763	0	0.31069	0.63586	0.00160887
4	0.001078	0	0.000173871	0.000104322	0.997774	0.000869354
5	0.214286	0	0	0	0.738095	0.047619
	0	1	2	3	4	5
0	0.992993	3.91435e-05	0.00140917	0.0027009	0.00285748	0
1	0.111958	0	0.750239	0.00159541	0.136168	3.98851e-05
2	0.794475	0	0.203266	0.000434329	0.00182418	0
3	0.0996038	0	0	0.854556	0.0458404	0
4	0.000305686	0	0	0.00906868	0.990626	0
5	0.115789	0	0	0	0.705263	0.178947

FIGURE 9: Matrices de confusion des trois tests avec 30% des données en apprentissage après fusion des positions couchées.

Cette fusion marque une amélioration de la méthode pour deux des trois truies, et une amélioration nette de la prédiction de la position couchée pour la truie 3. La position genoux est un autre sujet de discussion. En effet, elle est très mal prédite, notamment dû au fait que c'est une position peu occupée par la truie. C'est cependant une position importante, car c'est une position de transition. Une truie debout va se mettre sur ses genoux pour s'allonger plus doucement. Cela traduit donc un caractère précautionneux de la truie qu'il est important de repérer. S'il était donc tentant de fusionner la position genoux avec la position debout ou couché, agir ainsi serait contre-productif pour l'étude du comportement de la truie.

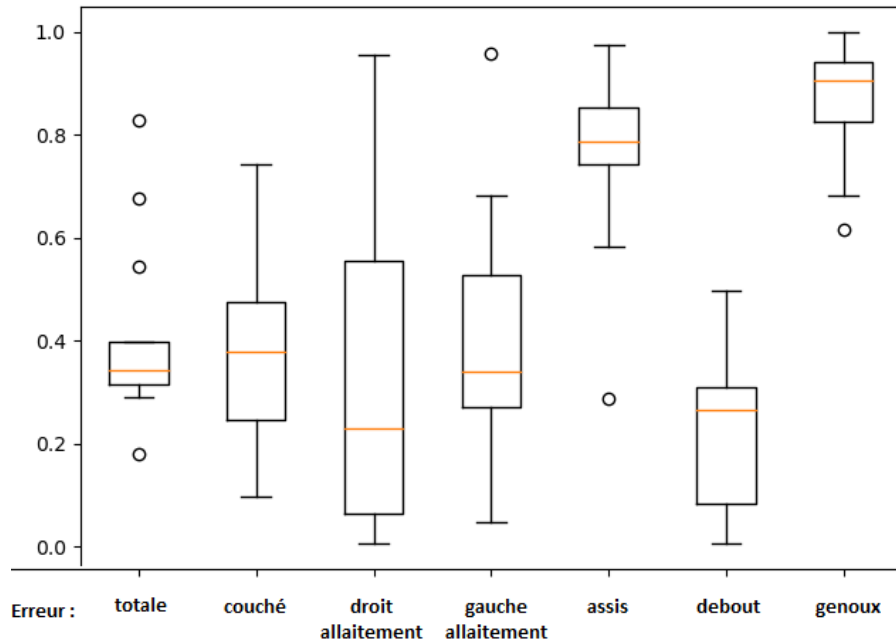


FIGURE 10: Diagramme en boîte représentant l'erreur de classification dans les tests avec toutes les truies sauf une comme entraînement pour chaque position

3.6. Un réseau pré-entraîné pour toutes les truies

L'objectif idéal serait d'établir une méthode de classification pré-entraîné qui fonctionnerait sur toutes les truies filmées dans les mêmes conditions. L'idée serait donc d'entraîner le réseau sur suffisamment de truies pour pouvoir prédire efficacement n'importe quelle truie. Au cours du stage, plusieurs tests sur ce genre de réseaux furent faits avec des résultats généralement mauvais. Voici le test le plus complet fait dans cette optique en fin de stage, sur les 13 truies disponibles.

Afin de tester au mieux l'effet d'un grand jeu d'entraînement sur la prédiction, on entraîne le réseau sur toutes les truies sauf une. On répète l'opération pour toutes les truies.

L'erreur moyenne obtenue sur ces tests est de 0.4068. La figure 10 montre la répartition de l'erreur totale pour chaque truie, mais aussi de l'erreur dans chaque position. La figure 11 montre la matrice de confusion moyenne de ces tests.

3.7. Un réseau pour une truie

Puisqu'un réseau global n'est pas une option viable, on en revient à créer un réseau par truie. Cela revient donc, pour chaque truie qu'on voudra étudier, à noter à la main une partie de son activité afin de pouvoir prédire le reste. Étant donné qu'annoter les vidéos est une activité coûteuse en temps pour un

	0	1	2	3	4	5
0	0.615772	0.0808242	0.155223	0.0685303	0.0727433	0.00690756
1	0.172396	0.626221	0.127838	0.00589105	0.0662844	0.00136937
2	0.220385	0.0384651	0.673236	0.00334238	0.0583817	0.00618967
3	0.553262	0.0108984	0.0124664	0.24479	0.178414	0.000169909
4	0.157202	0.00708969	0.0241564	0.0749447	0.734004	0.00260241
5	0.305708	0.0253823	0.0857288	0.0485219	0.402956	0.131702

FIGURE 11: Moyenne des matrices de confusion des tests, normalisée selon les lignes. 0 : couché. 1 : allaitement droit. 2 : allaitement gauche. 3 : assis. 4 : debout. 5 : genoux

expert, l'enjeu sera de déterminer quelle quantité d'information il est nécessaire d'annoter pour avoir une bonne prédiction.

Références

- Lambertz, C., Petig, M., Elkmann, A., Gaulty, M., 2015. Confinement of sows for different periods during lactation : effects on behaviour and lesions of sows and performance of piglets. *animal* 9, 1373–1378.
- Liu, D., Yu, J., 2009. Otsu method and k-means, in : 2009 Ninth International Conference on Hybrid Intelligent Systems, IEEE. pp. 344–349.
- Pedersen, J.H., 2015. Loose housing or temporary confinement of sows in designed farrowing pens. SL Grafik.
- Singh, C., Verdon, M., Cronin, G., Hemsworth, P., 2017. The behaviour and welfare of sows and piglets in farrowing crates or lactation pens. *animal* 11, 1210–1221.