

Maschinelle Lernverfahren zur Prognose von Tierwohrlisiken in der Schweinehaltung

Tobias Zimpel, Martin Riekert, Achim Klein, Christa Hoffmann

Tierwohl ist ein Qualitätsmerkmal moderner Schweinehaltung und zunehmend im Blickpunkt der Öffentlichkeit. Tierwohrlisiken sind multifaktoriell und müssen erkannt werden, bevor das Tierwohl gefährdet ist. Diese Arbeit nutzt Maschinelles Lernen (ML) zur frühzeitigen Prognose von Tierwohrlisiken. Der verwendete Datensatz umfasst Daten von über 57.000 Schweinen, gegliedert nach 10 Tierwohrlisiken und 14 Indikatoren für die Säugephase. Der zentrale Beitrag ist ein ML-Modell zur Prognose von Todesfällen in der Säugephase mit einer Accuracy von 80,4 %. Die Accuracy des Majority-Vote-Klassifikators für den Todesfall in der Säugephase beträgt hingegen nur 53,1 %. Somit könnte die Methode dazu beitragen, drohende Todesfälle in der Säugephase von Schweinen frühzeitig zu erkennen und Maßnahmen zu ergreifen.

Schlüsselwörter

Tierwohl, geburtsnahe Tierwohrlisiken, Schweinehaltung, Maschinelles Lernen

Tierwohl kann über die fünf Freiheiten („Five Freedoms“) nach dem FARM ANIMAL WELFARE COUNCIL (2009) definiert werden: die Freiheit von Hunger, Durst, Unbehagen, Schmerzen, Ängsten und Leiden, Verletzungen und Krankheiten sowie die Freiheit zur Entfaltung normaler Verhaltensweisen. Tierwohrlisiken werden in der Praxis beispielsweise durch Beobachtung von Tierwohlintikatoren erkannt (ZAPF et al. 2017). Ablauf und Hinweise zur Erhebung von Tierwohlintikatoren in der Praxis sind unter anderem in den Leitfäden des Kuratoriums für Technik und Bauwesen in der Landwirtschaft e.V. enthalten.

In der Literatur wird zudem der Einsatz von Sensoren beschrieben, um Indikatoren für das Tierwohl zu überwachen. So werden u. a. Kameras und Mikrofone verwendet (RIEKERT et al. 2020, MATTHEWS et al. 2016). Mikrofone konnten bereits zur Bestimmung des Stress-Levels eingesetzt werden. Der Klassifikationsfehler des Verfahrens betrug weniger als 5 % (MANTEUFFEL und SCHÖN 2002). CHUNG et al. (2013) berichten beim Einsatz von Mikrofonen zur automatischen Erkennung von Krankheiten von einem Klassifikationsfehler von 6 bis 9%. Sensorbasierte Ansätze zur Überwachung von Tierwohrlisiken setzen allerdings die Installation und den Betrieb von Sensoren voraus.

Ein anderer Ansatz liegt in der Nutzung von Daten als Indikatoren, die im Betrieb bereits erfasst werden, z. B. das Gewicht. Erste Vorarbeiten zur Prognose von Tierwohrlisiken weisen auf die Bedeutung von frühzeitigen biologischen Indikatoren hin (DÍAZ et al. 2017). So untersuchten DÍAZ et al. (2017) Zusammenhänge zwischen frühzeitig erfassten biologischen Indikatoren und der Sterblichkeit sowie der Gesundheit von Schweinen. Die Wahrscheinlichkeit eines Todesfalls von Schweinen mit einem Geburtsgewicht kleiner 0,95 kg liegt nach DÍAZ et al. (2017) bei 28,4%.

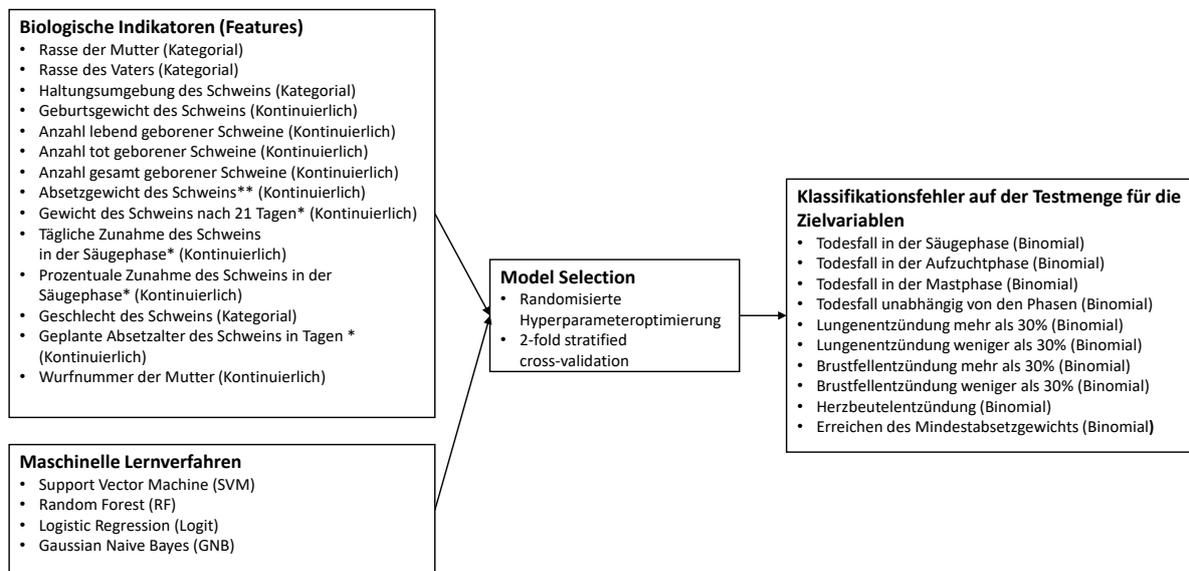
Schweine mit einem Geburtsgewicht größer 0,95 kg überlebten nach DÍAZ et al. (2017) zu 87,1%. Allerdings ist das Verfahren aufgrund der niedrigen Untergrenze sehr restriktiv. Weitere von DÍAZ et al. ermittelten Zusammenhänge zwischen frühzeitigen biologischen Indikatoren und Tierwohrlisiken sind Gegenstand der vorliegenden Arbeit.

siken beziehen sich auf Erkrankungen und Entzündungen von Organen. Demnach haben Schweine mit niedrigeren Absetzgewichten ein höheres Risiko für Herzerkrankungen, Lungen- und Brustfellentzündungen als Schweine mit höheren Absetzgewichten. Die Erkenntnis des höheren Risikos für Herzerkrankungen bei niedrigem Absetzgewicht lässt sich, zumindest für den Bluthochdruck als Herzerkrankung, auf ein niedriges Geburtsgewicht übertragen (POORE et al. 2002). POORE et al. (2002) konstatierten Schweinen mit einem Geburtsgewicht von weniger als 1,4 kg ein höheres Risiko zur Erkrankung an Bluthochdruck als Schweinen mit einem Geburtsgewicht größer 1,7 kg.

Vor diesem Hintergrund setzt diese Arbeit maschinelle Lernverfahren (ML) ein, um Tierwohrlisiken mit biologischen Indikatoren möglichst frühzeitig zu prognostizieren. Ein ML ist ein Algorithmus, der seine Leistung mit zunehmender Erfahrung hinsichtlich der Aufgabe steigert (MITCHELL 1997). In dieser Arbeit werden Support Vector Machine (SVM) (CORTES und VAPNIK 1995), Random Forest (RF) (BREIMAN 2001), Gaussian Naive Bayes (GNB) (JOHN und LANGLEY 1995) und Logistic Regression (Logit) (LIN et al. 2008) genutzt. Einen wesentlichen Einfluss auf diese Algorithmen haben die Hyperparameter. Eine geeignete Wahl der Hyperparameter erhöht die Leistung der ML, weshalb unterschiedliche Belegungen der Hyperparameter getestet werden (BERGSTRA, YAMINS und COX 2013).

Material und Methode

Zur Auswertung wurde ein Datensatz der Landesanstalt für Schweinezucht Boxberg verwendet, der 57.104 Beispiele mit unterschiedlichen biologischen Daten (u. a. Gewichte, Haltungsformen, Krankheiten) von Schweinen enthält. Die biologischen Daten der Schweine wurden, gemessen am Geburtsdatum der Schweine, im Zeitraum von Oktober 2011 bis Mai 2018 erfasst. Das Vorgehen zur Entwicklung von Modellen zur Prognose von Tierwohrlisiken teilt sich in die Auswahl, die Parametrisierung der Verfahren und das Training der Verfahren auf der Trainingsmenge und die Berechnung der Leistungsfähigkeit auf einer unabhängigen Testmenge auf. Abbildung 1 zeigt das Forschungsmodell dieser Arbeit mit 10 verschiedenen Tierwohrlisiken als Zielvariablen.



** Nicht für den Todesfall in der Säugephase und das Erreichen des Mindestabsetzgewichts

* Nicht für den Todesfall in der Säugephase

Abbildung 1: Forschungsmodell

Diese 10 Tierwohrlisiken umfassen 4 Ausprägungen an Todesfällen, 5 Erkrankungen und das Erreichen des Mindestabsetzgewichts. Das Mindestabsetzgewicht gilt in dieser Arbeit als erreicht, wenn das Schwein am Absetztermin über ein Gewicht von mindestens 5 kg verfügt. Die vier Ausprägungen der Todesfälle sind der Todesfall in der Säugephase, der Todesfall in der Aufzuchtphase, der Todesfall in der Mastphase und der Todesfall unabhängig von den Phasen der Schweinezucht. Die fünf Erkrankungen sind: die Lungenentzündung mit einem Krankheitsgrad von weniger als 30% der Lunge, die Lungenentzündung mit einem Krankheitsgrad von mehr als 30% der Lunge, die Herzbeutelentzündung, die Brustfellentzündung mit einem Schweregrad der Krankheit von weniger als 30% der Gewebeschicht und die Brustfellentzündung, die mehr als 30% der Gewebeschicht betrifft. Diese 10 Tierwohrlisiken wurden gewählt, da sie einen starken Einfluss auf die Entwicklung der Schweine haben (BOTREAU et al. 2007). Jedes Tierwohrlisiko ist binär kodiert mit dem Wert 1 für „tritt ein“ und 0 für „tritt nicht ein“. In den Daten sind die Krankheiten hinterlegt, wenn diese eingetreten sind. Daher wird angenommen, dass kein Eintrag der Krankheit auch bedeutet, dass diese Krankheit nicht eingetreten ist. Die Todesfälle werden dahingehend kategorisiert, ob ein Todesfall oder eine reguläre Schlachtung vorliegt.

Als Merkmale wurden 14 biologische Indikatoren ausgewählt, die in der Säugephase verfügbar sind und unmittelbar mit dem Schwein in Zusammenhang stehen (Abbildung 1):

1. die Haltungsumgebung des Schweins während der Säugephase, Aufzucht und Mast mit den Ausprägungen alternative und konventionelle Haltung
2. 8 Mutterrassen
3. 7 Vatterassen
4. das Geburtsgewicht jedes Tieres
5. das Kontrollgewicht am 21. Tag nach der Geburt
6. das Absetzgewicht
7. das Geschlecht mit den Ausprägungen männlich, weiblich und Zwitter

8. die prozentuale Zunahme zwischen Absetzgewicht und Geburtsgewicht
9. das geplante Absetzalter in Tagen
10. die tägliche Zunahme in der Säugephase
11. die Wurfnummer der Mutter (Anzahl an Würfen)
12. die Anzahl der lebend geborenen Schweine im gleichen Wurf
13. die Anzahl der tot geborenen Schweine im gleichen Wurf
14. die Anzahl an insgesamt geborenen Schweine im gleichen Wurf

Die Haltungsumgebung des Schweins beschreibt entweder die konventionelle Haltung auf einem Spaltenboden oder die alternative Haltung, die sich hinsichtlich Bodenausführung und Stallfläche je Tier unterscheidet (BUNDESMINISTERIUM FÜR ERNÄHRUNG UND LANDWIRTSCHAFT 2018, MAUER 2014). Die 8 Rassen der Mutter (überwiegend Kreuzung aus Large White und der Deutschen Landrasse sowie Kreuzung aus Large White, Leicoma und Deutsche Landrasse) und die 7 Rassen des Vaters (überwiegend Deutsches Piétrainschwein) im Datensatz entsprechen der Rassencodierung des Bundesverbands Rind und Schwein e.V.

Zur Datenebereinigung wurden die Schritte Identifizierung von Ausreißern, Transformation in Schreibweise bei Textfeldern und Überprüfung des Datentyps durchgeführt. Die Bereinigung von Ausreißern wurde nur für kontinuierliche Features durchgeführt, d. h. biologische Indikatoren mit einem Wert aus den reellen Zahlen (z. B. Geburtsgewicht). Features sind die Eingabedaten in ein ML. Als Ausreißer wurden Werte identifiziert, die größer als die Obergrenze des 1,5-fachen Interquartilsabstands plus 5% des Interquartilsabstands zum dritten Quartil (Q3) oder kleiner als die Untergrenze des 1,5-fachen Interquartilsabstands minus 5% des Interquartilsabstands zum ersten Quartil (Q1) sind. Diese Ausreißer wurden aus dem Datensatz vollständig entfernt. Beispiele wurden aus dem gesamten Datensatz entfernt, wenn der Ist-Datentyp des Einflussfaktors nicht mit dem Soll-Datentyp übereinstimmte. Schweine, bei denen ein oder mehrere Werte fehlten, wurden ebenfalls aus dem Datensatz entfernt. Alle Todesfälle, die nicht den letzten Wert als Todesfall oder Schlachtung aufwiesen, wurden aus dem Datensatz entfernt. Das Erreichen des Mindestabsetzgewichts war gegeben, wenn das Absetzgewicht größer gleich 5 kg war.

Für jedes Tierwohlrisiko wurde ein eigener Datensatz erstellt. Denn wenn z. B. Schweine in der Säugephase gestorben sind, können diese nicht im Datensatz für den Todesfall in der Mastphase vorkommen. Insgesamt wurden 10 Datensätze mit einer Größe von bis zu 11.000 Beispielen erstellt. Hierzu wurden die nach dem Geburtsdatum aktuellsten Beispiele gewählt. Jeder der Datensätze wurde vor dem One Hot Encoding in eine Trainings- und Testmenge aufgeteilt. Hierbei wurden je 10.000 Beispiele für die Trainingsmenge verwendet und die restlichen Beispiele als Testmenge. One Hot Encoding beschreibt das Ersetzen eines kategorialen Features durch Einfügen einer neuen Spalte für jeden möglichen Wert eines kategorialen Features. Hier beschreiben kategoriale Features biologische Indikatoren, bei denen die Werte nominal skaliert sind (z. B. Geschlecht). Nach der Aufteilung wurden Beispiele aus der Testmenge entfernt, bei denen Ausprägungen von kategorialen Features (z. B. Haltung „alternativ“) in der Testmenge vorkommen, jedoch nicht in der Trainingsmenge. Wird eine Ausprägung eines kategorialen Features in der Testmenge verwendet, die nicht in der Trainingsmenge enthalten ist, ändern sich die Bedeutung der Spalten und möglicherweise die Anzahl der Spalten nach dem One Hot Encoding. Die Anzahl der Beispiele in der Trainings- und Testmenge sind in Tabelle 1 dargestellt.

Tabelle 1: Anzahl an Beispielen in der Trainings- und Testmenge

Tierwohrlisiko	TF Säuge	TF Aufzucht	TF Mast	TF	HZ	BF1	BF2	LZ1	LZ2	MA
Beispiele in der Trainingsmenge	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000
Beispiele in der Testmenge	652	264	999	428	1.000	1.000	1.000	1.000	652	71

BF1: Brustfellentzündung < 30%, BF2: Brustfellentzündung > 30%, HZ: Herzbeutelentzündung, LZ1: Lungenentzündung < 30%, LZ2: Lungenentzündung > 30%, MA: Erreichen des Mindestabsetzgewichts, TF: Todesfall unabhängig von den Phasen, TF Aufzucht: Todesfall in der Aufzuchtphase, TF Mast: Todesfall in der Mastphase, TF Säuge: Todesfall in der Säugephase

In dieser Arbeit wurden die Algorithmen SVM (Support Vector Machine), RF (Random Forest) GNB (Gaussian Naive Bayes) und Logit (Logistic Regression, Scikit-learn Version 0.19.2) angewendet. Da kein Verfahren a priori eine höhere Leistungsfähigkeit aufweist, ist ein Vergleich der Verfahren erforderlich (WOLPERT 1996). Der Programmcode ist unter <https://wi2.uni-hohenheim.de/analytics> zur Replikation der Ergebnisse veröffentlicht (PINEAU et al. 2020).

Die Trainingsdaten wurden jeweils vor der randomisierten Hyperparameteroptimierung für jedes Tierwohrlisiko auf das Intervall [0,1] normalisiert und mit dem One-Hot-Encoding-Verfahren für kategoriale Features in Vektoren transformiert (BERGSTRA und BENGIO 2012). Diese Normalisierung ermöglicht eine kürzere Berechnungszeit durch Beibehaltung von Null-Werten im Vergleich zu Nicht-Null-Werten bei kategorialen Features. Mit derselben Normalisierungseinstellung der jeweiligen Trainingsmenge wurde auch die entsprechende Testmenge normalisiert.

Die Model Selection erfolgt durch randomisierte Hyperparameteroptimierung. Hierbei werden unterschiedliche Konfigurationen an Hyperparametern (Tabelle 2) auf der jeweiligen Trainingsmenge getestet, um die Parameter mit der höchsten Leistung des jeweiligen Verfahrens zu ermitteln (BERGSTRA und BENGIO 2012).

Tabelle 2: Werte für die randomisierte Hyperparameteroptimierung

ML	Hyperparameter	Wertebereich
RF	Anzahl Features für den Split	Zufälliger ganzzahliger Wert zwischen 1 und der Anzahl an Features
	Untergrenze an Beispielen für einen Split	Zufälliger ganzzahliger Wert zwischen 2 und 11
	Bootstrap	Wahr, Falsch
	Splitkriterium	Gini-Index, Entropie
	Anzahl Bäume	10, 100, 200
SVM	Kernel	linear, poly, rbf, sigmoid
	C	0,001, 0,01, 0,1, 1, 10, 50, 75, 100, 125, 150, 1.000, 1.500, 2.000, 3.000, 5.000
Logit	C	0,001, 0,01, 0,1, 1, 10, 50, 75, 100, 125, 150, 1.000, 1.500, 2.000, 3.000, 5.000

Für das Verfahren RF wurde für die Anzahl an Features eine zufällige natürliche Zahl zwischen 1 und der Anzahl an biologischen Indikatoren und für die minimale Anzahl an Stichproben eine zufällige natürliche Zahl zwischen 1 und 11 genutzt. Weiter wurde der gesamte Datensatz zur Baumerzeugung, für den Split die Verfahren Gini-Index und Entropie und für die Anzahl an Bäumen die Zahlen 10, 100 und 200 genutzt. Beim Verfahren Gini-Index werden die relativen Häufigkeiten der Feature-Werte als Entscheidungsgrundlage für den Split verwendet. Entropie beschreibt hingegen ein Maß für den Informationsgehalt, der beim Split als Entscheidungsgrundlage genutzt wird.

Bei SVM und Logit stehen in Anlehnung an Hsu et al. (2003), die Werte 0,0001, 0,01, 0,1, 1, 10, 50, 75, 100, 125, 150, 1.000, 1.500, 2.000, 3.000 und 5.000 zur Auswahl. Für den Hyperparameter „kernel“ stehen bei SVM ein linearer Kernel (linear), ein polynomialer Kernel (poly), ein Kernel mit radialer Basisfunktion (rbf) und ein Kernel mit Sigmoidfunktion (sigmoid) zur Auswahl. Bei der randomisierten Hyperoptimierung wurden für die genutzten ML jeweils zehn Durchgänge mit zufälligen Parametern der vordefinierten Parametermenge mit einer 2-fach stratifizierten Kreuzvalidierung durchgeführt (KOHAVI 1995). Beim GNB wurde keine Hyperparameteroptimierung durchgeführt, da keine Hyperparameter des Verfahrens belegt werden können, die das Verfahren nicht eingrenzen (wie z. B. Kosten der SVM).

Für jedes Tierwohlrisiko wurde zusammen mit der Hyperparameteroptimierung die Leistung aller Kombinationen an biologischen Indikatoren für das Training der Verfahren miteinander verglichen (z. B. Geburtsgewicht, Absatzgewicht und die beiden Merkmale Geburtsgewicht und Absatzgewicht zusammen). Somit ergeben sich bei 13 Features jeweils über 16.000 Kombinationen für die Tierwohlrisiken Todesfall in der Aufzuchtphase und in der Mastphase sowie für die Erkrankungen. Bei 9 Features des Todesfalls in der Säugephase und unabhängig von den Phasen wurden für beide Variablen jeweils über 500 Kombinationen verglichen. Für das Erreichen des Mindestabsatzgewichts mit 10 Features wurden hingegen über 4.000 Kombinationen verglichen.

Die Accuracy beschreibt, mit welchem prozentualen Anteil ein Klassifikationsverfahren Objekte einer Klasse korrekt zuordnet (WITTEN et al. 2011, NGUYEN und ARMITAGE 2008). Die Klassifikation des ML wird in folgende Kategorien eingeteilt (MANNING und SCHÜTZE 1999, S. 374):

- True positive (tp): Die Klasse wurde erkannt und die Klasse liegt vor.
- False positive (fp): Die Klasse wurde erkannt, allerdings ist diese Klasse nicht zutreffend.
- False negative (fn): Die Klasse wurde nicht erkannt, allerdings liegt die Klasse vor.
- True negative (tn): Die Klasse wurde nicht erkannt und liegt nicht vor.

Unter Verwendung der Konfusionsmatrix-Terminologie wurden hierbei die folgenden Metriken berechnet (MANNING und SCHÜTZE 1999):

$$Recall = \frac{tp}{tp + fn}$$

$$Precision = \frac{tp}{tp + fp}$$

$$Accuracy = \frac{tp + fp}{tp + fp + tn + fn}$$

Der Majority Vote ist eine Entscheidungsregel, bei der die am häufigsten vorkommende Klasse als wahre Klasse ausgewählt wird. Der Majority Vote wird folgend berechnet (TIAN und ZHU 2015):

$$\hat{y}_i = \operatorname{argmax}_{d \in [D]} \sum_{j=1}^N \mathbb{1}(x_{ij} = d), \forall i \in [M]$$

Wobei \hat{y}_i die ausgewählte Klasse für ein Beispiel i eines Datensatz mit M Beispielen, $[D]$ die Menge an vorkommenden Klassen (hier: $[D] = \{0, 1\}$), j ein Verfahren der N Verfahren (hier: $|N|=1$) für Vorschläge der Klasse und x_{ij} die vorgeschlagene Klasse von j für i darstellt. $\mathbb{1}(\cdot)$ ist eine Indikatorfunktion, die den Wert 1 annimmt, wenn die vorgeschlagene Klasse der tatsächlichen Klasse entspricht und ansonsten den Wert 0 annimmt (TIAN und ZHU 2015).

Zum Vergleich aller Kombinationen an biologischen Indikatoren bei der Parametrisierung und dem Training der Verfahren der einzelnen Tierwohrlisiken wird die Accuracy genutzt. Für die trainierten ML-Verfahren der Kombinationen an biologischen Indikatoren mit der höchsten Accuracy auf der Validierungsmenge eines Tierwohrlisikos wurden die Metriken Recall, Precision sowie Accuracy auf der unabhängigen Testmenge berechnet.

Ergebnisse

In Tabelle 3 ist die Accuracy für die Tierwohrlisiken nach der Hyperparameteroptimierung in der Model Selection mit 2-fach stratifizierter Kreuzvalidierung dargestellt. Für den Todesfall unabhängig von den Phasen wird auf der Trainingsmenge eine Accuracy von 0,736 und für den Todesfall in der Säugephase eine Accuracy von 0,837 erreicht. Auf der Validierungsmenge wurden keine Precision und kein Recall berechnet. Für die Erkrankungen, bestehend aus den Lungenentzündungen, der Herzbeutelentzündung und den Brustfellentzündungen, dem Erreichen des Mindestabsetzgewichts, dem Todesfall in der Aufzuchtphase sowie dem Todesfall in der Mastphase wird keine Verbesserung der Accuracy in Relation zum Majority Vote der Klassen in der Model Selection erzielt, die größer als 0,02 ist. Für die restlichen Tierwohrlisiken sind die Konfigurationen nicht angegeben, da in der Model Selection und auf der Testmenge keine Verbesserung der Accuracy erzielt wird.

Tabelle 3: Accuracy in der Model Selection

	TF	TF Säuge	MA	TF Aufzucht	TF Mast	HZ	BF1	BF2	LZ1	LZ2
Majority Vote	0,502	0,717	0,980	0,951	0,984	0,980	0,983	0,998	0,929	0,993
SVM	0,736	0,836	0,996	0,951	0,984	0,980	0,983	0,998	0,929	0,993
RF	0,696	0,830	0,993	0,951	0,984	0,981	0,983	0,999	0,929	0,993
Logit	0,729	0,837	0,994	0,951	0,984	0,980	0,983	0,998	0,929	0,993
GNB	0,733	0,663	0,966	0,934	0,984	0,980	0,983	0,998	0,925	0,993

BF1: Brustfellentzündung < 30 %, BF2: Brustfellentzündung > 30 %, GNB: Gaussian Naive Bayes, HZ: Herzbeutelentzündung, LZ1: Lungenentzündung < 30 %, LZ2: Lungenentzündung > 30 %, MA: Erreichen des Mindestabsetzgewichts, Logit: Logistic Regression, RF: Random Forest, SVM: Support Vector Machine, TF: Todesfall unabhängig von den Phasen, TF Aufzucht: Todesfall in der Aufzuchtphase, TF Mast: Todesfall in der Mastphase, TF Säuge: Todesfall in der Säugephase

Für jedes Tierwohrlisiko wurde das ML-Modell ausgewählt, das die höchste Accuracy aufwies und auf der jeweiligen Testmenge getestet. So wurde z. B. die SVM als ML-Modell für den Todesfall unabhän-

gig von den Phasen und die Logit als ML-Modell für den Todesfall in der Säugephase ausgewählt. Die damit erzielte Accuracy, Precision und Recall auf der Testmenge sind in Tabelle 4 dargestellt.

Tabelle 4: Metriken des ausgewählten ML-Verfahrens auf der Testmenge

	TF	TF Säuge	MA**	TF Aufzucht*	TF Mast*	HZ*	BF1*	BF2*	LZ1*	LZ2*
Accuracy des Majority Vote	0,661	0,531	0,972	0,909	0,990	0,981	0,992	0,995	0,956	0,988
Accuracy des gewählten ML-Modells	0,755	0,804	0,986	0,909	0,990	0,981	0,992	0,995	0,956	0,988
Precision des gewählten ML-Modells	0,964	0,807	0,986	0,807	0	0	0	0	0	0
Recall des gewählten ML-Modells	0,682	0,765	1	0,765	0	0	0	0	0	0
Anzahl Beispiele der Testmenge	428	652	71	652	999	428	1.000	1.000	1.000	652

** Testmenge kleiner als 1 % der Validierungsmenge

* Keine Verbesserung gegenüber Majority Vote

BF1: Brustfellentzündung < 30 %, BF2: Brustfellentzündung > 30 %, GNB: Gaussian Naive Bayes, HZ: Herzbeutelentzündung, LZ1: Lungenentzündung < 30 %, LZ2: Lungenentzündung > 30 %, MA: Erreichen des Mindestabsetzgewichts, ML: Maschinelles Lernverfahren, Logit: Logistic Regression, RF: Random Forest, SVM: Support Vector Machine, TF: Todesfall unabhängig von den Phasen, TF Aufzucht: Todesfall in der Aufzuchtphase, TF Mast: Todesfall in der Mastphase, TF Säuge: Todesfall in der Säugephase

Der Wert 0 für Precision und Recall in Tabelle 4 bezieht sich darauf, dass es für das ML nicht möglich ist, die Klasse zu unterscheiden und dass es daher den Majority Vote für „tritt nicht ein“ verwendet. Die Hyperparameter des ML-Modells für den Todesfall in der Säugephase, den Todesfall unabhängig von den Phasen und das Erreichen des Mindestabsetzgewichts sind in Tabelle 5 dargestellt.

Tabelle 5: Konfiguration des aus der Modellauswahl hervorgehenden ML-Verfahrens

	TF	TF Säuge	MA
Biologische Indikatoren	AggS, GS, GW, W	AlgS, AtgS, AggS, GW, GS, H, RM, RV, W	AlgS, AT, GW, H, TZN
Gewähltes ML-Verfahren	SVM	Logit	SVM
Hyperparameter des gewählten ML-Verfahrens	C: 0.1 kernel: rbf	C: 1	C: 5.000 kernel: rbf

AggS: Anzahl der insgesamt geborenen Schweine, AlgS: Anzahl lebend geborener Schweine, AT: Absetzalter in Tagen, AtgS: Anzahl tot geborener Schweine, GS: Geschlecht, GW: Geburtsgewicht, H: Haltung des Schweins, Logit: Logistic Regression, MA: Erreichen des Mindestabsetzgewichts, RM: Rasse der Mutter, RV: Rasse des Vaters, SVM: Support Vector Machine, TF: Todesfall unabhängig von den Phasen, TF Aufzucht: Todesfall in der Aufzuchtphase, TF Mast: Todesfall in der Mastphase, TF Säuge: Todesfall in der Säugephase, TZN: Tägliche Zunahme in der Säugephase, W: Wurfnummer

Für den Todesfall in der Säugephase wird eine Erhöhung der Accuracy von 0,531 auf 0,804 ermöglicht. Diese Erhöhung geht einher mit einer Precision von 0,807 und dem Recall von 0,765. Die Erhöhung der Accuracy um 0,273 entspricht einer Erhöhung der Accuracy von über 51,41 % gegenüber der Nutzung des Majority Votes zur Klassifikation. Die Accuracy des Todesfalls unabhängig von den Phasen wird beim Majority Vote von 0,661 um 0,094 auf 0,755 mit dem ML-Modell erhöht. Diese Differenz von 0,094 entspricht einer Erhöhung der Accuracy um über 14,22 %. Hierbei wird beim Todesfall unabhängig von den Phasen eine Precision von 0,964 und ein Recall von 0,682 erreicht.

Für die Erkrankungen, den Todesfall in der Aufzuchtphase, den Todesfall in der Mastphase wird keine Verbesserung der Accuracy in Relation zum Majority Vote auf der Testmenge erzielt. Für das Erreichen des Mindestabsetzgewichts wird eine Verbesserung der Accuracy gegenüber dem Majority Vote erreicht, allerdings liegt die Größe der Testmenge bei weniger als 1 % der Trainingsmenge. Für die Tierwohrlisiken ohne Verbesserung gegenüber dem Majority Vote ist die Konfiguration nicht angegeben, da ein Einsatz des ML-Modells nicht notwendig ist, um dieselbe Accuracy zu erreichen.

Als ML-Modell für den Todesfall unabhängig von den Phasen wurde die SVM mit dem Wert 0,1 für den Hyperparameter „C“ und „rbf“ als Kernel genutzt. Für den Todesfall in der Säugephase hingegen wurde die Logit mit dem Wert 1 für den Hyperparameter „C“ verwendet. Als biologische Indikatoren wurden für den Todesfall unabhängig von den Phasen die Gesamtzahl an geborenen Schweinen, das Geschlecht, das Geburtsgewicht und die Wurfnummer verwendet. Für den Todesfall in der Säugephase wurde die Anzahl lebend geborener Schweine, die Anzahl tot geborener Schweine, die Gesamtzahl an geborenen Schweinen, das Geburtsgewicht, das Geschlecht, die Haltungsform, die Rasse der Mutter, die Rasse des Vaters und die Wurfnummer als biologische Indikatoren genutzt. Das Geschlecht, die Wurfnummer, die Gesamtzahl an geborenen Schweinen sowie die Rasse der Mutter sind in den beiden genutzten Kombinationen an biologischen Indikatoren enthalten, welche die größte Erhöhung der Accuracy über die verschiedenen Tierwohrlisiken hinweg erzielen.

Diskussion

Der Beitrag dieser Arbeit ist ein ML-Modell zur Prognose des Todesfalls in der Säugephase und ein Modell zur Prognose des Todesfalls unabhängig von den Phasen in der Schweinezucht. Die relative Steigerung der Accuracy gegenüber der dem Majority Vote beträgt für den Todesfall in der Säugephase 51,41 % und für den Todesfall unabhängig von den Phasen 14,22 %. Die genutzten biologischen Indikatoren in den ML-Modellen beschreiben keinen Erklärungszusammenhang zwischen den Indikatoren und den Tierwohrlisiken. DÍAZ et al. (2007) beschreiben den Einfluss des Geschlechts und der Rasse der Mutter auf Tierwohrlisiken. Im ML-Modell zum Todesfall in der Säugephase werden ebenfalls das Geschlecht und die Rasse der Mutter als Indikatoren verwendet (Tabelle 5). Die verwendeten Indikatoren Wurfnummer und Gesamtzahl an geborenen Schweinen beschreiben möglicherweise den Einfluss vorheriger Würfe, mit Blick auf das Alter der Sau (Wurfnummer) oder die Größe des aktuellen Wurfs. So ist es denkbar, dass die Gesamtzahl an geborenen Schweinen das Geburtsgewicht eines Schweins beeinflusst. Der Einfluss des Geburtsgewichts auf den Todesfall in der Säugephase ist in der Literatur bekannt (DÍAZ et al. 2017).

Für den Todesfall in der Aufzucht- und Mastphase, die Krankheiten und das Erreichen des Mindestabsetzgewichts konnte kein ML-Modell erzeugt werden, das gegenüber dem Majority Vote eine Erhöhung der Accuracy erreicht. Eine Erklärung könnte sein, dass diese Risiken durch komplexere Einflüsse bedingt sind, die in den genutzten Daten nicht entsprechend repräsentiert waren oder dass Indikatoren fehlen, die erst gegen Mitte bis Ende des Lebenszyklus erfasst werden. Dies folgt der Beobachtung, dass die Accuracy für den Todesfall unabhängig von den Phasen niedriger ist als die Accuracy für den Todesfall in der Säugephase. Der Todesfall unabhängig von den Phasen stellt eine Generalisierung des Todesfalls in den verschiedenen Phasen dar. So unterscheiden sich auch die Hyperparameter der genutzten Modelle auf der Trainingsmenge zwischen dem Todesfall in der Säugephase und unabhängig von den Phasen.

Möglicherweise werden Todesfälle nach der Säugephase und Erkrankungen durch komplexere Zusammenhänge beeinflusst, als in den Daten der biologischen Indikatoren abgebildet, wodurch die Accuracy sich nicht erhöhte. Daraus ergeben sich Ansatzpunkte für die weitere Forschung. So ist die Verwendung weiterer biologischer Indikatoren denkbar, wie die Nutzung von Sensordaten (z. B. Wasserdurchflussmengen, Futtermengen oder Temperaturdaten). Ein weiterer Ansatz kann in der Veränderung der Hyperparameterwerte bei gleichem Datensatz liegen. Auch die Überprüfung der Übertragbarkeit der ML-Modelle auf andere Tierarten oder die Nutzung von Daten anderer Quellen ergeben Ansätze für weitere Forschung.

Diese Arbeit unterliegt folgenden Grenzen. Die Testmenge des Todesfalls in der Säugephase entspricht 4,28% der Trainingsmenge. Die Testmenge des Todesfalls unabhängig von den Phasen entspricht 6,52% der Trainingsmenge. Beide Testmengen liegen damit unter den anvisierten 10% oder 1.000 Beispielen. Eine größere Testmenge würde die Aussagekraft über die Prognosefähigkeit bei neuen Beispielen erhöhen. Die Prozentsätze der Testmenge, die niedriger als anvisiert waren, entstanden durch die Aufteilung des gesamten Datensatzes in eine Trainings- und eine Testmenge *vor* dem One Hot Encoding. Hierdurch wurden auf der einen Seite deskriptive Statistiken sowohl für die Trainings- als auch die Testmenge erzeugt, auf der anderen Seite führte dies zu Klassen in Features in den Beispielen der Testmenge, die nicht in der Trainingsmenge enthalten sind und daher entfernt wurden. Bei einer Aufteilung *nach* dem One Hot Encoding wären keine Beispiele aus diesem Grund entfernt worden.

Eine weitere Einschränkung ergibt sich durch die unausgeglichene Klassenhäufigkeit der Zielvariablen. Dies führt dazu, dass die Verfahren auf eine Vorhersage des Majority Vote optimiert sind, da die Klassifikation des Majority Vote seltener zu einem Fehler führt. Hierdurch erreichen die Verfahren die höchste Accuracy, allerdings kann im betrieblichen Kontext ein hoher Recall der Tierwohrlisiken erwünscht sein, um diese frühzeitig verhindern zu können. Hierzu existieren weitere Verfahren, die zukünftig in der Forschung über Tierwohrlisiken untersucht werden können (HAIXIANG et al. 2017).

Anstatt eines Datensatzes für alle Tierwohrlisiken wurde für jedes Tierwohrlisiko ein eigener Datensatz verwendet. Dies führt zu unterschiedlichen Beispielen von Schweinen in den einzelnen Datensätzen. So wird verhindert, dass ein Schwein, welches in der Säugephase gestorben ist, im Datensatz des Todesfalls der Aufzuchtphase oder Mast nochmals vorkommt. Ebenso wurden in der Trainings- und Testmenge die bis zu 11.000 aktuellsten Beispiele genutzt und nicht zufällig ausgewählt, in der Annahme, dass die Umstände bei der Erfassung (z. B. die Buchtenausstattung) näher an den tatsächlichen aktuellen Umständen liegen und sich dies positiv auf die Prognosefähigkeit auswirkt. Allerdings sind weitere Umstände, wie z. B. Epidemien oder Futterumstellungen, nicht im Datensatz berücksichtigt und können die Prognosefähigkeit beeinträchtigen.

Die genutzte Min-Max-Normalisierung gewährleistet einen effizienten Einsatz der ML beim One Hot Encoding, allerdings ist die Min-Max-Normalisierung anfällig für Extremwerte. Daher wurde jeder kontinuierliche Indikator auf Ausreißer überprüft. Bei einer Standardisierung mit der z-Transformation würde die Seltenheit der Nicht-Null-Features hingegen nicht erhalten bleiben, weshalb die Min-Max-Normalisierung genutzt wurde. Die z-Transformation beschreibt die Umwandlung von Features, sodass diese einen Erwartungswert von null und eine Varianz von eins haben. Speziell durch das One Hot Encoding der Rasse der Mutter und der Rasse des Vaters im Zusammenspiel mit den anderen biologischen Indikatoren und der Anzahl an verschiedenen Tierwohrlisiken war der

effiziente Einsatz der ML – im Hinblick auf die benötigte Berechnungsdauer bei der Erstellung eines ML-Modells bei gegebener Rechenleistung – notwendig.

Schlussfolgerungen

Diese Arbeit beschreibt ein Modell der Logistic Regression, das die Accuracy für den Todesfalls in der Säugephase gegenüber dem Majority Vote von 53,1 auf 80,4 % erhöht. Diese Erhöhung geht einher mit einer Precision von 0,807 und einem Recall von 0,765. Für den Todesfall unabhängig von den Phasen wird ein Modell der Support Vector Machine beschrieben, das eine Erhöhung der Accuracy gegenüber dem Majority Vote von 66,1 auf 75,5 % ermöglicht. Hierbei werden eine Precision von 0,964 und ein Recall von 0,682 erzielt. Zur Prognose des Todesfalls in der Säugephase wurden biologische Indikatoren verwendet, die bereits bei der Geburt verfügbar sind und ohne spezielle Sensoren ermittelt werden. Somit kann in der Praxis geburtsnah eine Prognose des Todesfalls durchgeführt werden, um gegebenenfalls Gegenmaßnahmen durchzuführen.

Literatur

- Bergstra, J.; Bengio Y. (2012): Random Search for Hyper-Parameter Optimization., Journal of Machine Learning Research 13, pp. 281-305
- Bergstra, J.; Yamins, D.; & Cox, D.D. (2013): Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. Proceedings of the 12th Python in science conference 13, pp. 13-19, <https://doi.org/10.25080/Majora-8b375195-003>
- Botreau, R.; Veissier, I.; Butterworth, A.; Bracke M.B.M.; Keeling, L.J. (2007): Definition of criteria for overall assessment of animal welfare. Animal Welfare 16, pp. 225-228
- Breiman, L. (2001): Random Forests. Machine Learning 45, pp. 5-32
- Bundesministerium für Ernährung und Landwirtschaft (2018): Landwirtschaft verstehen - Fakten und Hintergründe. www.bmel.de/SharedDocs/Downloads/Broschueren/Landwirtschaft-verstehen.pdf;jsessionid=4672B911DF00C56AF288442A5EA518E0.1_cid367?__blob=publicationFile, Zugriff am 23.12.2018
- Bundesverband Rind und Schwein e. V. (2020): LPA-Rassencodes. <https://www.rind-schwein.de/brs-schwein/lpa-rassencodes.html>, Zugriff am 17.04.2020
- Chung, Y.; Oh, S.; Lee, J.; Park, D.; Chang, H-H; Kim, S. (2013): Automatic Detection and Recognition of Pig Wasting Diseases Using Sound Data in Audio Surveillance Systems. Sensors 13(10), pp. 12929-12942
- Cortes, C.; Vapnik, V. (1995): Support-Vector Networks. Machine Learning 20, pp. 273-297
- Díaz, J. A. C.; Boyle, L. A.; Diana, A.; Leonard, F. C.; Moriarty, J. P.; McElroy, M. C.; McGettrick, S.; Kelliher, D.; Manzanilla, E. G. (2017): Early life indicators predict mortality, illness, reduced welfare and carcass characteristics in finisher pigs. Preventive Veterinary Medicine 146, pp. 94-102, <https://doi.org/10.1016/j.prevetmed.2017.07.018>
- Farm Animal Welfare Council (2009): Farm Animal Welfare in Great Britain: Past, Present and Future. <https://www.gov.uk/government/publications/fawc-report-on-farm-animal-welfare-in-great-britain-past-present-and-future>, Zugriff am 17.04.2020
- Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. (2017): Learning from class-imbalanced data: Review of methods and applications, Expert Systems with Applications 73, pp. 220-239, <https://doi.org/10.1016/j.eswa.2016.12.035>
- Hsu, C.-W.; C.-C. Chang; Lin, C.-J. (2003): A Practical Guide to Support Vector Classification, Technischer Bericht, National Taiwan University
- John, G. H.; Langley, P. (1995): Estimating continuous distributions in Bayesian classifiers, in: Hg. Besnard, P.; Hanks, S., Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montréal Qué, Morgan Kaufmann Publishers, 1. Auflage., pp. 338-345

- Kohavi, R. (1995): A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence (2), pp. 1137-1143
- Lin, C.-J.; Wenig, R. C.; Keerthi, S. S. (2008): Trust Region Newton Method for Large-Scale Logistic Regression. Journal of Machine Learning Research 9, pp. 627-650
- Manning, C. D.; Schütze, H.; (1999): Foundations of Statistical Natural Language Processing. Cambridge, The MIT Press
- Manteuffel, G.; Schön, P.-Ch. (2002): Measuring welfare of pigs by automatic monitoring of stress sounds. Measurement Systems for Animal Data - Bornimer Agrartechnische Berichte 29, pp. 110-118
- Matthews, S. G.; Miller, A. L.; Clapp, J.; Plötz, T.; Kyriazakis, I. (2016): Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. The Veterinary Journal 217, pp. 43-51, <https://doi.org/10.1016/j.tvjl.2016.09.005>
- Mauer, J. (2014): Vergleich der Schweinemast in Stallungen der konventionellen und alternativen Bauweise. www.lsz-bw.de/pb/site/lel/get/documents/MLR.LEL/PB5Documents/lsz/pdf/Fachinformationen/Schweinemast%20-%20Haltung/LSZ_Vergleich%20Schweinemast%20in%20Stallungen.pdf?attachment=true, Zugriff am 23.12.2018
- Mitchell, T. M. (1997): Machine Learning. New York, The McGraw-Hill Companies
- Nguyen, T. T.; Armitage, G. (2008): A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys & Tutorials 10(4), pp. 56-76, <https://doi.org/10.1109/SURV.2008.080406>
- Pineau, J.; Vincent-Lamarre, P.; Sinha, K.; Larivière, V.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; Larochelle, H. (2020): Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program). arXiv preprint, arXiv:2003.12206
- Poore, K. R.; Forhead, A. J.; Gardner, D. S.; Giussani, D. A.; Fowden, A. L. (2002): The effects of birth weight on basal cardiovascular function in pigs at 3 months of age. Journal of Physiology 539(3), pp. 969-978, <http://dx.doi.org/10.1013/jphysiol.2001.012926>
- Riekert, M.; Klein, A.; Adrion, F.; Hoffmann, C.; Gallmann, E. (2020): Automatically detecting pig position and posture by 2D camera imaging and deep learning. Computers and Electronics in Agriculture 174, p. 105391, <https://doi.org/10.1016/j.compag.2020.105391>
- Tian, T. und Zhu, J. (2015): Max-Margin Majority Voting for Learning from Crowds. Advances in neural information processing systems, pp. 1621-1629
- Witten, I. H.; Frank, E.; Hall, M. A. (2011): Data Mining: Practical Machine Learning Tools and Techniques. Amsterdam, Morgan Kaufmann Publishers, 3. Auflage
- Wolpert, D. H. (1996): The Lack of A Priori Distinctions Between Learning Algorithms. Neural Computation 8(7), pp. 1341-1390
- Zapf, R.; Schultheiß, U.; Knierim, U.; Brinkmann, J.; Schrader, L. (2017): Assessing farm animal welfare – guidelines for on-farm self-assessment, Landtechnik 72(4), pp. 214-220, <https://doi.org/10.1515/lt.2017.3166>

Autoren

M. Sc. Tobias Zimpel, **M. Sc. Martin Riekert** und **Dr. oec. Achim Klein** sind wissenschaftliche Mitarbeiter am Lehrstuhl für Wirtschaftsinformatik der Universität Hohenheim, Schwercstraße 35, 70599 Stuttgart.
E-Mail: tobias.zimpel@uni-hohenheim.de

Dr. Christa Hoffmann war während der Durchführung der Studie Referatsleiterin an der Landesanstalt für Schweinezucht Boxberg, Seehöfer Str. 50, 97944 Boxberg-Windischbuch

Danksagung

Die Autoren danken den Mitarbeitern der Landesanstalt für Schweinezucht LSZ für ihre Unterstützung und Beratung während der Experimente und Bereitstellung des Datensatzes. Besonderer Dank geht an Andrea Wild, Günter Lenkner, Lilljana Pyrkotsch und Hansjörg Schrader.

Das Projekt „Landwirtschaft 4.0: Info-System“ wird finanziert aus Geldern des Ministeriums für Ländlichen Raum und Verbraucherschutz Baden-Württemberg.