www.ThePharmaJournal.com

The Pharma Innovation

**ISSN (E): 2277-7695**
**ISSN (P): 2349-8242**
**NAAS Rating: 5.23**
TPI 2022; SP-11(9): 2829-2843

www.thepharmajournal.com
Received: 22-06-2022
Accepted: 26-07-2022

**Anuradha Panwar**
Ph.D. Scholar, Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Harshit Kumar**
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Divya Rajawat**
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Sonali Sonejita Nayak**
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Kanika Ghildiyal**
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Anurodh Sharma**
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Atul Singh Rajput**
Livestock Production and Management Section, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Gangula Lokavya Reddy**
Division of Animal Reproduction, Indian Council of Agricultural Research (ICAR) -Indian Veterinary Research Institute (IVRI), Izatnagar, Uttar Pradesh, India

**Manjit Panigrahi**
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

**Corresponding Author:**
**Anuradha Panwar**
Ph.D. Scholar, Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

# High throughput phenotyping and big data analytics for livestock improvement

**Anuradha Panwar, Harshit Kumar, Divya Rajawat, Sonali Sonejita Nayak, Kanika Ghildiyal, Anurodh Sharma, Atul Singh Rajput, Gangula Lokavya Reddy and Manjit Panigrahi**

## Abstract
Tremendous progress has been made continually with the over-expanding genomics technologies to uncover and understand animal genomes. However, the impact of genomics data on animal improvement is still far from satisfactory, largely due to the lack of effective phenotypic data; our capacity to collect useful high-quality phenotypic data lags behind the current capacity to generate high-throughput genomics data. Thus, the research bottleneck in animal sciences is shifting from genotyping to phenotyping. High-throughput phenotyping techniques offer a new opportunity to enhance genomic improvement of livestock, especially for novel phenotypes. Together the growing demand for food and the advancement in sensing technology has the potential to make animal farming more centralized, large-scale, and efficient. The use of sensors, big data, artificial intelligence, and machine learning can help animal farmers to lower production costs, increase efficiencies, enhance animal welfare and grow more animals per hectare. One of the most relevant challenges in this context is the handling of large-scale data provided by automated processes such as image collection, continuous real-time sensor-based measurements, and spectroscopy reports, among others. The extraction of biologically relevant features from large datasets generated by automatic devices can be done further by using machine learning algorithms. Many studies have demonstrated the usefulness of advanced remote sensing technologies coupled with machine learning (ML) approaches for the accurate prediction of valuable animal traits. Although AI and ML algorithms have developed so fast, there is a lack of standardization in the collection and sharing of data globally. However, as more farms get connected to technology, AI and sensing technologies will start playing a more decisive role in helping farmers see patterns and solutions to pressing problems in modern animal farming.

**Keywords:** Artificial intelligence, high-throughput phenotyping, machine-learning, sensors

## Introduction
Enhancing milk and meat productivity has long been a global concern. Researchers are realizing that current breeding projects will not be sufficient to meet anticipated future food demands under the present climate scenario. Screening animals that lead to overall yield in the field necessitates immediate action. Nowadays, traditional breeding operations are being transformed into more efficient modern breeding programs by incorporating emerging technologies, most notably high-throughput phenotyping (Crossa *et al*., 2021) [23]. Although genomic information on various animal breeds is publicly available and accessible online, the phenotypic data on their genomes is still limited, as environmental factors impede phenome characterization and increase the likelihood of error in the measurements of traits (Rahaman *et al*. 2015) [92]. Genomic selection (GS) is an approach that uses genome-wide marker information to predict genomic estimated breeding values of lines in a breeding population (Meuwissen *et al*. 2001) [71]. When compared to other traditional approaches, such as marker-assisted selection (MAS), GS has some inherent advantages, including increasing genetic gain by shortening breeding cycles (Heffner *et al*. 2010) [45] and capturing minor effect loci based on markers spread across the entire target genome (Hayes *et al*. 2009) [43]. Accurate prediction model training for GS requires reliable phenotypes in addition to genotyping. Due to high labor and time costs, phenotyping is considered an important factor limiting genetic gains in animal breeding. As a result, attempts have been made to develop high-throughput phenotyping (HTP) platforms because traditional phenotyping techniques are prohibitively expensive, time-consuming, slow, and frequently harmful, and they can only analyse a few variables at a time (Hein *et al*., 2021) [46].

Accurate and large-scale phenotypic data are required for successful animal breeding programs and for the genomic dissection of complicated traits. Non-destructive phenotyping, a modern technology, provides an additional dimension for data collection by increasing the speed, precision, and analysis of captured data (Yang *et al*., 2020) [124]. Digital image analysis extracts meaningful information from images and can be used as an input for imaging processing techniques that directly apply to livestock phenotyping. Other advancements in technology, such as automatic feeding systems, activity monitor sensors, and indirect biomarkers at physiological and cellular levels, have the potential to provide a wide range of novel phenotypes. Furthermore, infrared spectrometry is gaining popularity in precision livestock farming as a non-destructive measurement tool and a valuable resource for online analysis. Historically, animal farming has always been decentralized, on a scale that a few people can manage. And until about a century ago, most animal farmers lacked access to advance technologies like high-speed internet, smartphones, and low-cost computing power. Both of these conditions are rapidly changing right now. HTP is essential as the global demand for various meat and animal products is expected to rise by more than 70% over the next three decades (Rojas-Downing *et al*., 2017) [95] and secondly, more than half of the world's population is directly linked to the internet via smartphones or computers. A large number of animal farmers now have easy access to computing power. We now know that meat consumption has increased wherever populations and incomes have increased. Because land and other natural resources are limited, we will need to identify more effective ways to grow more animals per hectare to meet this growing demand and we must now produce animals with fewer resources such as water, land, and other natural resources. This also implies that manual animal farming processes may no longer be adequate. It also means that we must devise methods and systems to help us achieve greater profits in animal farming. Computers, sensors, cloud computing, machine learning (ML), and artificial intelligence (AI) are already transforming various industries. They generate greater profits and efficiencies (Wolfert *et al*., 2017) [121]. This is why we must investigate how advanced technologies can assist us in achieving greater efficiencies and gains in animal farming. Brito *et al*. argue that using large-scale phenotyping to evaluate genomic traits for animal welfare-related traits is a better solution for accurate selection in a commercial production system (Rajawat *et al*., 2022a; Rajawat *et al*., 2022b) [83, 94]. They described the main bioinformatic and statistical tools available for this purpose, with a focus on approaches to developing novel HTP-based welfare indicator traits, such as movement recording using accelerometers and wearable sensors. Sensor-based phenotypes, like IMU-based movement capture, infrared thermography, and sound analysis, in conjunction with big data science, are critical for translating animal welfare indicators traits into precise genomic breeding values. This can be used in commercial selective breeding to continue improving animal resilience. The ability to collect and use on-farm data for breeding purposes transformed the beef, dairy, poultry, and swine industries, resulting in massive productivity and efficiency gains (Hill, 2016) [49]. Similar opportunities for enormous profits may exist in several other contexts of the livestock industry using sensors and other high-throughput phenotyping technologies (Science Breakthroughs to Advance Food and Agricultural Research by 2030, 2019).

**Sources of significant expenses in livestock farming**

Stocking rate, feeding, and disease management can all be major cost drivers in animal farming. Farmers can optimise their major expenses and decrease their production costs by increasing the number of animals in a system due to economies of scale (Rojas-Downing *et al*., 2017) [95]. Feed intake and efficiency are essential contributors to sustainability in the dairy cattle industry because they have an economic and environmental impact. The feed has the most significant economic impact on dairy farm profitability, accounting for more than 40% of milk production expenses (National Milk Cost of Production). Feed efficiency improvements will also benefit the environment by lowering greenhouse gas emissions from cattle and manure, as well as land requirements for manure disposal and water needs (Knowlton *et al*., 2004; Von Keyserlingk *et al*., 2013) [55, 117]. In a large animal farm, where thousands of animals are housed together, a contagious disease outbreak can result in significant losses. The infectious disease outbreak will be difficult to contain in such a setting unless the farmer takes timely early interventions. When symptoms appear, it is frequently too late to intervene. If a disease is allowed to spread unchecked, it will result in animal deaths, poorer health outcomes, and financial losses. On the other hand, an intelligent farm with multiple sensors may alert the farmer to abnormal animal behavior in many initial phases.

**HTP Platforms**

Over the last decade, a variety of high-throughput phenotyping methods/platforms have been used in animals for the identification of functional limiting factors, determining the optimal nutrient composition of animal feed, evaluating animal management to evaluate performance (Ferguson *et al*., 2014) [31], examining strategies for lowering nutrient excretion into the environment (Pomar *et al*., 2019) [90], or forecasting outcomes in other covariates (Ferguson *et al*., 2014) [31]. Animal phenotyping is constantly evolving, with low throughput phenotyping and invasive methods being overtaken by non-destructive, high-throughput methods (Rahaman *et al*., 2015; Mir *et al*., 2015) [92, 73]. Over the last decade, rapid advances in non-destructive inexpensive sensors and imaging techniques have revolutionized animal phenomics. The current non-destructive high-throughput phenotyping platform (Figure 1) uses sophisticated technologies such as activity sensors such as accelerometers, pedometers, and GPS devices to track the location, speed, and time of movement. A sensor method is used to predict calving and lameness, and an activity-based ODS can be used to manually select cows in oestrus. The use of non-invasive sensors can enable the phenotyping of several thousand of animals in a day, similar to the impact of high throughput DNA sequencing technology in animal genomics (Finkel *et al*., 2009) [33]. (Table 1, shows the various studies done by using phenomics platforms for trait phenotyping in animals). It does, however, entail generating an unprecedented amount of complex data. Data storage and transfer can be limiting, especially in the field. Furthermore, image processing is a time-consuming task that limits the ability of these tools to achieve high-throughput screening. Same is true for managing high-throughput data of animal breeding (Saravanan *et al*., 2019) [99].

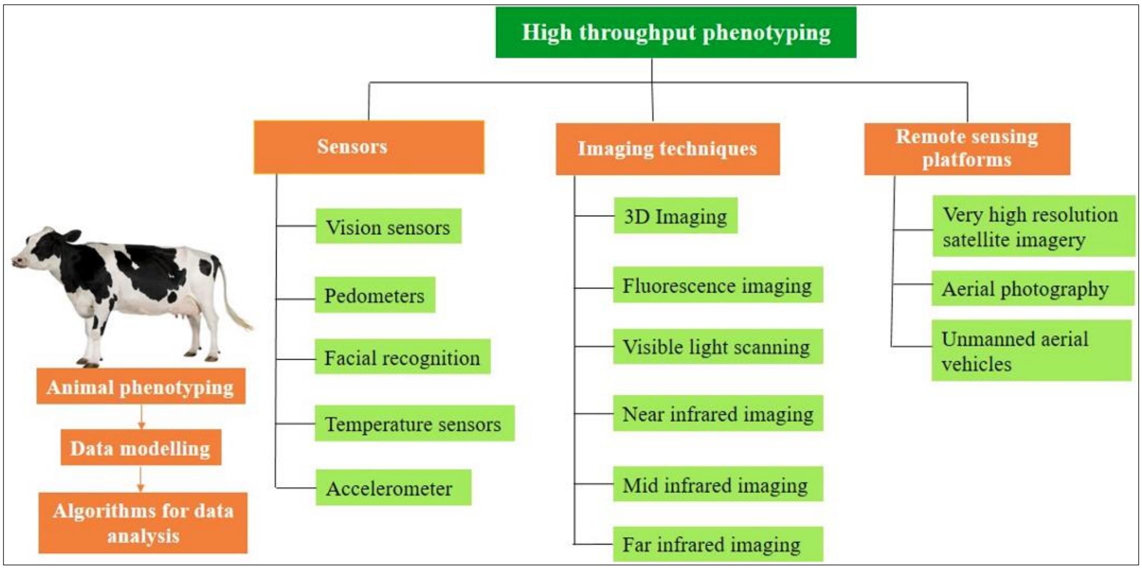


**Fig 1:** Schematic representation of High throughput Phenotyping in Animals

**Table 1:** Studies conducted by using various phenomics platforms for trait phenotyping in animals

| Platform/Recording technique | Trait studied | Method applied | Reference |
|---|---|---|---|
| Remotely sensed imagery (Very high-resolution satellite imagery, aerial photography and UAVs) | Cows posture ('standing', 'lying' and 'grazing) | Deep learning methods (Nanonets API, ENVI 5.4 deep learning module, and YOLO v3) | Mucher *et al*., 2022 [129] |
| Ultra-wide band positioning system | Lying behavior | Bagged tree algorithm | Adriaens *et al*., 2022 [130] |
| Sensor | Resilience trait of dairy cows | Random forest algorithms | Ouweltjes *et al*., 2021 [81] |
| Sensor | Lameness | Additive logistic regression | Kamphuis *et al*., 2013 [131] |
| Sensor | Calving management | Logistic regression models | Rutten *et al*., 2017 [96] |
| Kinect Camera | Muscles grading in pigs | Gradient boosted classifier | Alsahaf *et al*., 2019 [7] |

## Phenotyping platforms in trait phenotyping: some examples

By making use of cutting-edge technologies, many issues in animal farming can be resolved. Finding the best options to reduce expenses, raise output, boost efficiency, and develop the best diet formulations are a few examples (Ferguson *et al*., 2014) [31]. Advanced models may even take into account variables like genetics, environment, and management priorities in order to provide relevant and contextually optimal solutions. In general, the more diverse the datasets collected and analyzed by a system, the greater its chances of arriving at accurate and optimal solutions (Ellis *et al*., 2020) [29]. HTP can help in precise livestock farming in various ways, a few examples are discussed in this review.

## Monitoring animal health

High-throughput methods have also been used to continuously monitor key animal health parameters such as movement, air quality, and food and fluid consumption. By continuously gathering this data and utilising cutting-edge AI and ML algorithms to forecast deviations or irregularities, farmers can now identify, predict, and prevent disease outbreaks even before a significant attack (Neethirajan *et al*., 2020) [78]. A system like this can reduce production costs while also alerting farmers to the possibility of disease even during the preclinical stage. This, in turn, will assist farmers in taking timely action to avoid catastrophic losses (VanderWaal *et al*., 2017) [116]. At a much lower cost, they can immediately predict and stop the spread of contagious diseases like the African swine flu. The transmission of many infectious diseases may now be predicted by improved technology before it spreads widely, which is more significant.

## Lameness prediction in early stages

In other situations, algorithms can forecast disease symptoms like lameness based on the motions of the animal. Preclinical lameness can be consistently detected by modifications in gait, excessive use of certain body parts, and inactivity in other body parts (Taneja *et al*., 2020) [111]. Lameness is the third most important disease affecting farming (Neethiranjan *et al*., 2020) [78] because it reduces milk production and increases injury risk (Warner *et al*., 2020) [118]. Predicting lameness ahead of time can help farmers avoid significant financial losses. Another study found that infected pigs move less, by up to 10%, during the first two days of infection. This can be used to isolate infected animals before they infect a large number of other animals (Fernández-Carrión *et al*., 2017) [32]. Finally, farmers can help to prevent diarrhea and bacterial infections in pigs by using sensors that gather environmental data such as temperature, gas production, and humidity.

## Locating animals in an extensive production system

HTP has been used to locate animals in an extensive system, and information about cattle in extensive production systems can be obtained using remotely sensed imagery (such as very high-resolution satellite imagery, aerial photography, and unmanned aerial vehicles (UAVs)) and Convolutional Neural Networks (CNNs) as AI technology for object-based detection of cattle in large datasets (Kellenberger, Marcos and Tuia 2018) [54]. Using UAV imagery to identify individual cattle and their poses can provide information on resilience and efficiency.

**Monitoring of lying behavior of animals**

The lying behavior of animals can be observed by using ultra-wideband (uwb) positioning systems. It has been demonstrated that it changes in response to changes in health and welfare status (Tucker *et al*., 2021) [114]. Lameness, for example, reduces the number of times an animal gets up or lies down while increasing overall lying duration. Similarly, udder infections that cause an animal to become extremely ill and metabolic issues that affect rumination time will alter lying behavior. Accurate detection and monitoring of lying over time can reveal health and welfare status, contribute to new precision phenotypes, and accurately and non-invasively evaluate, for example, housing situations or management practices. One method is to use 3-dimensional spatial data, such as that provided by modern ultra-wideband positioning systems currently being developed and commercialized. Lameness, mastitis, and infertility have been identified as the top three dairy cow health issues associated with economic losses in the dairy industry (Juarez *et al*., 2003; Panigrahi *et al*., 2022a) [51, 93]. Lameness has a negative impact on welfare because it is associated with pain (Whay *et al*., 1997; Bicalho *et al*., 2007) [120, 11], and it reduces farm profitability due to poorer reproductive performance, loss of milk production, and increased treatment and culling costs (Tranter and Morris, 1991; Sprecher *et al*., 1997; Green *et al*., 2002) [112, 109, 38]. Typically, lame cows are identified visually by observing their stride and back position (Sprecher *et al*., 1997) [109]; however, in larger herds, as the number of cows maintained per farm labour unit rises, visual identification of lame cows becomes more challenging.

**Resilience detection in animals**

Lifetime resilient dairy cows are defined as animals with a high likelihood of completing multiple lactations, good productive and reproductive performance, few health problems that are easily overcome, and efficient and consistent milk production (Adriaens *et al*., 2020) [3]. Improving resilience in dairy cows has significant benefits: it enhances animal health and welfare (Mulder and Rashidi, 2017) [76], farm productivity (Colditz and Hine, 2016, Ouweltjes *et al*., 2021) [22], the need for antibiotics (König and May 2019) [58] and the sector's environmental impact. Recent technological advancements and increased use of other sensors provide the opportunity to combine continuous data recordings from various sensors to improve resilience prediction. In contrast to projections based purely on daily milk characteristics, Adriaens *et al*. (2020) [3] observed that incorporating activity sensor data considerably (p 0.01) enhanced resilience prediction accuracies. From a biological perspective, Poppe *et al*. (2020) [91] and Adriaens *et al*. (2020) [3] attempted to construct resilience indicators.

**Survival prediction in cattle**

Cow survival is a multifaceted trait that incorporates traits such as milk production, fertility, health, and environmental factors such as farm management (van der Heide *et al*., 2020) [115]. A high farm average of lactations obtained is another sign of successful farming methods with regard to animal care (Barkema *et al*., 2015) [10]. Because there are numerous advantages to cows that live long, productive lives, farmers would be wise to keep only those cows that are likely to thrive in a production environment (https://www.farmersweekly.co.za/). By predicting a cow's ability to survive early on, it would be possible to select cows with a high probability of surviving to higher lactations. The ensemble method (Knutti *et al*., 2010; Wozniak *et al*., 2014) [56, 123], also known as a hybrid classifier (Wozniak *et al*., 2014) [123], decision fusion method (Sinha *et al*., 2008) [108], or aggregation method, can be used to predict survival in dairy cattle (Satopaa *et al*., 2014) [105].

**Improving pig farming**

Several pig farms have made use of computer vision technology (Matthews *et al*., 2015) [67]. Attempts to analyze or estimate carcass composition in-vivo using image-based solutions are most relevant to our problem (Scholz *et al*., 2015; Carabs *et al*., 2016) [107, 17]. Kinect sensors have previously been used in pig farming applications such as monitoring and detecting pig behaviors in pen (Lee *et al*., 2016) [66], automated weight estimation (Kongsro *et al*., 2014; Pezzuolo *et al*., 2018) [57, 88], and walking pattern analysis (Stavrakakis *et al*., 2015) [110]. The ML algorithm can also be used to predict slaughter age in pigs, allowing for pig grouping prior to the initiation of the finishing phase (Alsahaf *et al*., 2018) [6]. Gradient Boosting Machine (GBM) models can be utilised to determine which ejaculate should be processed for AI doses (Kamphuis *et al*., 2020) [53]. Using ML algorithms such as the Gradient Boosting Machine algorithm, pigs that are prone to developing abnormal growth rates, meat percentages, or pneumonia during the growing-finishing phase can be identified; early detection would help to prevent heterogeneous pens through management (Mollenhorst *et al*., 2019) [74].

**Calving management in dairy cows**

Calving management is critical for the health and survival of dairy cows and their calves. A sensor system that predicts the moment of calving could help farmers check cows for calving more efficiently (Rutten *et al*., 2017) [96]. Observing a cow prior to calving is important because dystocia can occur, necessitating timely intervention to mitigate adverse effects on both the cow and the calf. Sensor data can be more valuable than the expected calving date alone in predicting the start of calving (Rutten *et al*., 2017) [96]. Sensors can automatically monitor several behavioral and physiological parameters associated with the advent of calving. Dairy cows' feeding and ruminating behavior decreases gradually in the last two weeks before calving (Büchel *et al*., 2014) [14] and drops abruptly at calving (Bar and Solomon, 2010) [9]. Sensors appear to be capable of identifying these variations (Bar and Solomon, 2010; Bucher and Sundrum, 2014; Schirmann *et al*., 2013) [9, 15, 106]. Time spent feeding decreases, as does dry matter intake (Schirmann *et al*., 2013; Bucher and Sundrum, 2014) [106, 15], and activity changes in the 24 hours preceding calving (Clark *et al*., 2015; Miedema *et al*., 2011; Saint-Dizier and Chastant-Maillard, 2015) [21, 72, 97]. Titler *et al*. (2015) [127] demonstrated that an activity index could accurately predict whether a cow would calve within 6 hours of an increment in the activity index. Previous research has shown that temperature (measured at the vulva, rectum, and rumen) drops during the 24 hours preceding calving (Saint Dizier and Chastant-Maillard, 2015) [97]. Ouellet *et al*. (2016) [80] proved that all of these sensor-measured parameters have predictive value for calving. If the start of calving could be predicted more accurately than the expected calving date, farmers would be able to identify when a cow requires intensive supervision.

### Detection of oestrus
Oestrus detection is regarded as one of the most labor-intensive and skilled tasks required of dairy farm personnel, and it is prone to a high error rate (McGowan *et al*., 2007; Hempstalk *et al*., 2010; Burke *et al*., 2012) [68, 47, 16]. Inexperienced farm staff, poor use of aids by farm staff, and weak oestrus displays by dairy cows are all factors that contribute to poor oestrus detection performance (Burke *et al*., 2012) [16]. Activity-based oestrus detection systems (ODS) are one such commercially available option that uses a pedometer or accelerometer technologies. These systems are based on the assumption that oestrus is associated with an increase in locomotion activity (Eradus *et al*. 1992) [30].

### Biomarker Development
Deep phenotyping can yield biomarkers that are sensitive, specific, and relatively inexpensive for detecting the trait or disease phenotype of interest for accurate animal classification. Any substance or process that can be measured in a biological specimen and is always associated with the trait of interest is a biomarker (Trusheim *et al*., 2007) [113]. Biomarkers can be molecules like RNA, metabolites, microbes, or proteins/peptides, but they can also be based on other modalities like imaging (Hartwell *et al*., 2006) [42]. Biomarker assays should ideally be minimally invasive, i.e., detectable in peripheral blood or urine (Hartwell *et al*., 2006; Ziegler *et al*., 2012) [42, 126]. Biomarkers are also used as screening tests for diseases that are subclinical, asymptomatic, or in their early stages. Furthermore, the benefits of early intervention/prevention (such as improved disease outcomes) should outweigh the costs of the screening test (Ziegler *et al*., 2012) [126]. Mid-infrared spectral data can be obtained quickly and cheaply from samples collected from monthly dairy herd improvement programs. A growing body of literature indicates that such data has excellent potential for predicting disease risk (Grelet *et al*., 2016) [39], greenhouse gas emissions (Dehareng *et al*., 2012) [25], and a variety of other physiological states in the cow (De Marchi *et al*., 2014; Gengler *et al*., 2016) [24, 34].

### Big data
To meet the needs of a growing human population, the livestock industry recognizes that there are challenges and opportunities to be more efficient, environmentally friendly, and societally conscious. These challenges present opportunities for more sustainable and profitable agriculture. The data types required to meet these grand challenges are diverse, presenting numerous opportunities for scientific discovery to link genotype to phenotype, develop computational tools for big data analytics, engineer new sensor technologies, develop new data coordination systems, and ultimately use this information for improved animal production and welfare (Panigrahi *et al*., 2020) [84]. The availability of low-cost computing power, massive storage media, and internet connectivity have exponentially increased the amount of data collected from individuals and groups of livestock. When the correct analytical framework is used, combining individual animal records such as weights, treatments, and carcass characteristics with cohort-level information such as daily feed deliveries, diet ingredients, group weights, and movements gives significant amounts of data useful for analysis. Automated systems for high-input data collection are becoming more common, resulting in an exponential increase in the availability of information on a wide range of hosts, pathogens, and environmental factors affecting animal health. For example, capturing whole genome sequences and gene expression data from both hosts and pathogens is becoming more common, as is the use of electronic medical records, animal genomics and digitized images (Ahmad *et al*., 2020; Chhotaray *et al*., 2020; Kaisa *et al*., 2020; Chhotaray *et al*., 2021a; Chhotaray *et al*., 2021b; Pal *et al*., 2022) [5, 18, 52, 19, 20, 82]. Genome sequences are also being utilized for detecting selection signatures and studying the nature of evolution in general (Saravanan *et al*., 2020a; Saravanan *et al*., 2020b; Saravanan *et al*., 2021a; Saravanan *et al*., 2021b) [101, 103, 100, 104]. Furthermore, sensors, drones, and satellites routinely collect data in huge amounts. Sensor technologies are increasingly being used to monitor animal welfare and health in cattle (Smith *et al*., 2006; Matthews *et al*., 2016) [67, 128], pigs (Guarino *et al*., 2017) [41], and poultry. This transition is pushing animal science into the era of Big Data, in which data sets are of lower fidelity and are collected quickly and in large quantities, and where equipment may vary significantly and in less coordinated ways (Parasar *et al*., 2021; Patra *et al*., 2021) [86, 87]. Because of the growing availability of such electronic data, massive databases have been created that are too big and complex to manage using traditional data analysis tools. Specialized tools are required in these cases to collect, organize, and analyze data. These datasets and the tools used to analyze them have been dubbed "big data." While the term "big data" has become ubiquitous, its meaning is frequently ambiguous, particularly because it tends to conflate data with the analytical methods used to analyze that data. Big data have many different definitions provided by authoritative sources (commercial technology organizations such as Microsoft and Google are frequently cited sources of definition), but big data "Vs" is a common theme amongst these. The so-called four V model (IBM, http://www.ibmbigdatahub.com/infographic/four-Vs-big-data) defines big data based on collection properties, but a literature review shows that the Vs are expanding at the same rate as the data itself, from 4- to 7- to 10- to 42-V models. Big data is defined by the four V models based on four key attributes: 1) volume, 2) velocity, 3) variety, and 4) veracity. Volume is simply the available amount of data. The speed with which users want to access or use data is called velocity. The various forms in which data is received are referred to as variety. Veracity focuses on the need to clean and edit large amounts of data in order to derive sound inferences from quality-controlled records. Value is frequently added to V models because, as data becomes less expensive to collect, the utility of those observations is often questioned unless there is an improvement in the methods/technology used to generate the data and, as a result, the quality of these data. However, it is unclear whether this is necessary in the livestock sector because the value proposition is central to the decision to collect data in the first place (Saravanan *et al*., 2022a) [102]. The use of "big data" will rely on careful data editing to remove noise and focus on the informative aspects of the data that are valuable for analytics, so quality validation is more important. The European Organization for Nuclear Research's Large Hadron Collider (ACM, 2011) and the Human Genome Project (Green *et al*., 2015) [37] are canonical examples of big data in the sciences (Saravanan *et al*., 2022b) [98]. Precision agriculture has a growing literature, which frequently overlaps with several big data concepts (e.g., Wolfert *et al*., 2017; Morota *et al*., 2018; Weersink *et al*., 2018; Mehrotra *et al*., 2021a; Mehrotra *et al*., 2021b) [122, 75, 119, 69, 70].

Precision agriculture aims to make appropriate management decisions by using detailed, frequently collected observations of individual animals. This includes identifying changes in productivity, determining reproductive status, early detection of health problems at the individual and group levels, and grouping animals with similar nutritional or other management needs in a livestock setting (Panigrahi *et al*., 2022b) [85].

The transition from the barn or pen level to the farm or landscape level (i.e., country, state, or regional location attributes) can be accompanied by a dramatic increase in the amount of data available, including soil composition, weather data, and water availability, and utilization. Apart from this it is also becoming increasingly important in animal identification at molecular level (Kumar *et al*., 2019: Kumar *et al*., 2021a; Kumar *et al*., 2021b; Kumar *et al*., 2021c; Kumar *et al*., 2021d; Kumar *et al*., 2022) [59, 60, 62, 63, 64, 61].

**Big data analytics**

Big data analytics is the process of analyzing large amounts of diverse data sets using advanced analytic techniques. These various data sets include unstructured, semi-structured, and structured data from multiple sources and sizes ranging from terabytes to zettabytes. Accelerometers and real-time location systems, for example, capture recursive numeric positions of individual animals in relatively short time steps. Algorithms can use this data to determine whether these movements are within or outside of expected ranges, as well as to identify patterns associated with physiologic changes such as estrus or disease (Table 2, shows how advanced algorithms can help animal farmers). Cohort and individual animal data, when combined with individual animal activity monitoring systems, can yield large data sets useful for both descriptive and predictive analytics. The methodology for using scientific information to make decisions is as follows: develop a hypothesis/prediction, test the prediction, evaluate the results, interpret the results, and revise and repeat the process (Larson and White, 2015) [65]. To discover meaningful patterns, predictive analytics employs various analytic techniques such as traditional statistics, machine learning, and data mining (Abbott, 2014) [1]. Predictive analytics is focused on making predictions, and these models can surpass traditional statistical models for several productions and research questions. An artificial neural network, a type of predictive regression model, was found to predict daily milk yields more accurately than traditional regression models in one study of livestock production data (Grzesiak *et al*., 2006) [40]. Another study used a variety of classification algorithms, including naive Bayesian classification, decision trees, random forests, and logistic regression, to predict the outcomes of diseased feedlot cattle. When evaluating overall accuracy, the authors discovered that logistic regression was rarely the best model (Amrine *et al*., 2014) [8]. Classification algorithms differ in their methodology or techniques for minimizing variation in outcomes of interest, and a growing library of potential machine learning algorithms is available for testing. Some models are better suited to different results or data types. Unlike statistical modeling, where a single model is chosen before analysis, multiple models can be tested, and the best algorithm is determined based on classification accuracy. The predictive analytic process includes selecting a target variable, handling the data, partitioning the data, creating algorithms, refining algorithms, and at last comparing the accuracy of the created classifiers. Each step in the process is critical to ensuring that the outcomes are internally valid and provide the information needed to improve subsequent decision-making Predictive analysis models use data, statistics, and machine learning techniques to improve farm effectiveness and drive successful outcomes in animal farming.

The cycle of the predictive analytics process can be conceptualized in a variety of ways, here Figure 2 is a list of some simple steps that can help in understanding and developing a successful predictive analysis. The predictive analytic process can include various steps like defining the problem statement, collecting data, cleaning data, partitioning data, analyzing data, creating a predictive model, validating it, deploying it, and monitoring it.

**Table 2:** How advanced algorithms can help animal farmers

| Trait of interest | Parameter detected | Algorithm used | References |
|---|---|---|---|
| Mastitis | Milk electrical conductivity (EC), other like milk colour, somatic cell count, and milk yield | Random Forest and Bayesian Network | Dhoble *et al*., 2019; Ebrahimi *et al*., 2019 [26, 28] |
| Lameness | Leg movement, Neck movement and Image/Video data | Fog computing, Classification and regressive tree (CART) XGBoost algorithm | Gertz *et al*., 2020; Taneja *et al*., 2020; Warner *et al*., 2020 [35, 111, 118] |
| African Swine Flu | Mobility, speed, direction | Optical flow algorithm | Fernandez-Carrión *et al*., 2017 [132] |
| Coccidiosis | Volatile Organic Compounds (VOC) in air | Principal Component Analysis (PCA) | Borgonovo *et al*., 2020 [12] |
| Postpartum disease | Lactose yield, Protein production, Milk yield | Random Forest Algorithm (RFA) | Hidalgo *et al*., 2018 [48] |
| Survival prediction | Milk production, fertility, and health | Ensemble method (logistic multiple regression, Naive Bayes and random forest) | Van der Heide *et al*., 2020 [115] |
| Nitrogen Excretion of Dairy Cattle | Milk production, milk composition, feed intake, and feed composition | Bayesian Network (BN) and Boosted regression trees (BRT) | Mollenhorst *et al*., 2020 [74] |
| Semen quality assessment in boar | Fertility and ejaculation record | Gradient Boosting Machine (GBM) models | Kamphuis *et al*., 2020 [53] |
| Growth rate and weight of pigs | Growth rate and weight gain | Ensemble methods: random forest, extremely randomized trees, gradient boosted machines and XGBoost | Alsahaf *et al*., 2018 [6] |

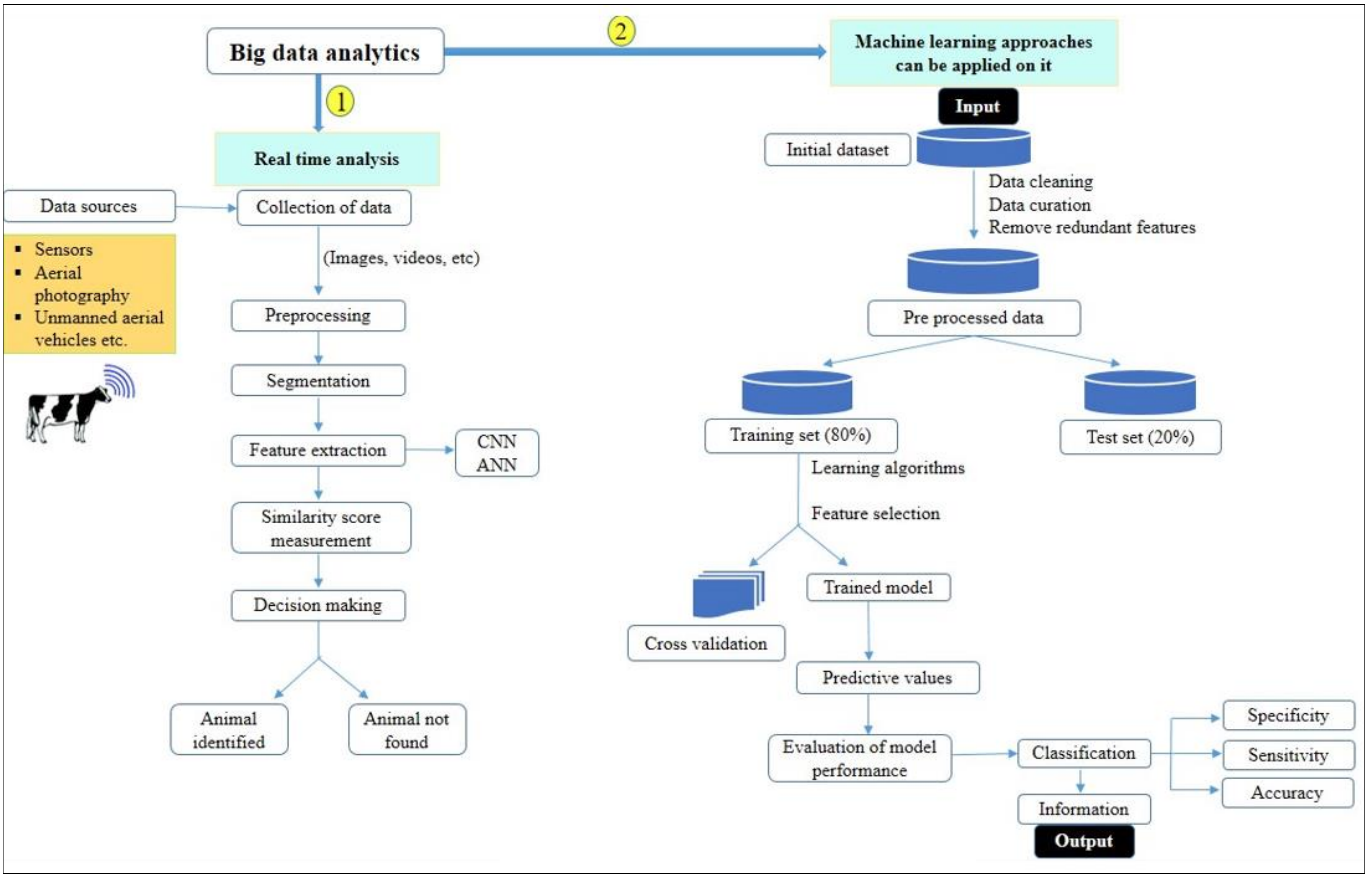


**Fig 2:** Basic work-flow of big data analytics

**Define the problem statement and data collection**

One of the first steps in the predictive analytic process is to identify the target variable, which is the outcome of the data to be estimated or predicted (Abbott, 2014) [1]. The target variable should be carefully chosen to provide information that will drive overall business decisions critical to the operation's financial sustainability. The structure of the predictive analytic problem and the appropriate model to be deployed is determined by the type of data describing the target variable. After selecting the target variable, the next step should be to clearly define the question to be answered. Narrowing the scope of a problem to an explicit hypothesis or question will inform the specific data required, limit the potential models, and define the level of accuracy that makes the model useful for decision-making (Abbott, 2014) [1]. A well-defined decision point will allow the model to focus on predicting a specific piece of information that can be used as a leverage point for operational decisions.

**Data cleaning/preprocessing**

Once the target variable has been determined, and basic data summarization has been used to evaluate the data structure, the data must be preprocessed. In general, the process entails assessing potential outliers, determining the level of missing data, and, if necessary, employing methods for evaluating collinearity among variables (crucial if using traditional methods like linear regression). Many machine learning algorithms handle missing data satisfactorily, which is one advantage of using these methods instead of conventional statistical methods when evaluating production data collected from livestock systems; however, large amounts of missing data can cause variables to become unreliable predictors, so missing data should be minimized if possible. Before beginning the modeling process, evaluate why the data are missing. This process is time-consuming, but it should be considered as should the most appropriate method to deal with the missing data (Abbott, 2014) [1]. Multicollinearity between two or more continuous variables can result in biased coefficient estimation and power loss (Yoo *et al.*, 2014) [125]. When using traditional methods such as multiple linear regression, Multicollinearity can lead to significant issues and should be strictly evaluated prior to model building. However, when using predictive methods such as random forests, collinear variables can be kept in the data set because they may be used in different bootstrapped trees, and each collinear variable may provide helpful information (Hayes *et al.*, 2015) [44].

**Data partitioning**

Following the completion of the variables to be included in the model, the available data is divided into training and validation/testing data sets. Prior to model building, data must be partitioned into these data subsets in order to evaluate predictive model performance and ensure that the model can be reasonably applied to new datasets. The first data subset is training data, which comprises approximately (80 percent). It is used to generate the initial models, and the training data subset may differ from the native data in terms of the frequency of occurrence of the outcome. It's time to put the model to the test after we've trained it with the training dataset. This dataset is used to assess the model's performance and ensure that it can generalise well to new or unfamiliar datasets. The test dataset is a subset of the original data that is distinct from the training dataset. It does, however, have some features and class probability distributions in common, and it is used as a benchmark for model evaluation once model training is complete. Test data is a well-organized dataset that contains data for each type of scenario for a given problem that the model would encounter if used in the real world. An ML project's test dataset is typically 20-25 percent of the total original data. The testing data should be representative of or a subset of the original dataset, and it should be large enough to provide meaningful predictions. There are no specific rules for the amount of data to partition into training, revising, and validation/testing datasets, but authors have successfully used partitions of 50% training, 25% revising, and 25% validation/test (Abell *et al*., 2017) [2] and 40% training, 30% revising, and 30% validation/test (Abell *et al*., 2017) [2]. (Amrine *et al*., 2014) [8]. The splits size is determined by the availability of data and the frequency with which outcomes occur, but the data have to be partitioned before the start of the predictive analytic process to avoid potential bias in the final predictive assessment.

**Model building**

Numerous types of predictive models are available to classify data or generate predictive outcomes based on a given data set. Big data contains many variables, many of which have complex interactions and relationships with the desired result; thus, a predefined model selection based on the target variable is frequently impossible. By evaluating the performance of multiple models, the evaluation does not assume that data will conform to a specific form, and the optimal model is determined by final accuracy evaluation rather than preconceived notions. The predictive analytic framework provides the environment for testing and evaluating multiple classifying algorithms to determine the best fit (in terms of accuracy) for a specific situation and target variable. Classification methods differ in their ability to handle missing data, deal with different attribute types (continuous or categorical), overall generalizability, and ability to provide a clear explanation for why a specific prediction was made. The target variable type should guide the particular algorithm chosen and the level of prediction accuracy required. A simple classification model is often an excellent place to start (e.g., a logistic regression model). The decision tree is another good initial model framework because of its ease of understanding and deployment in popular SQL-based database systems (Abbott, 2014) [1]. Both regression and classification problems can be solved using the decision tree framework. Compared to single-model methods, ensemble methods combine the results of multiple models to provide a more accurate predictive model. Random forests are a popular ensemble model that often outperforms single trees in terms of accuracy.

Multiple trees are generally constructed to predict the same target variable using different combinations of predictors and data set subsets. Using multiple methods, these various models are combined to determine the final expected outcome. The k-nearest neighbor (k-nn) method is one of the simplest models for predicting continuous outcomes. This model is easily described when using continuous variables to predict another continuous variable. The "k" in k-nn represents the number of animals used after calculating the distance. The more neighbors used to generate a prediction, the smoother the prediction; however, no theory specifies the number of neighbors to use (Abbott, 2014) [1]. Although the k-nn method is not as sophisticated as the other methods discussed, it can provide reasonably accurate results with minimal processing and is easily interpretable depending on the available target variable and predictor variables.

**Building predictive model**

Following the preprocessing steps, the next step is to train predictive models. Individual algorithms are given the required parameters, the target variable is specified, and the model is trained using the partitioned training data subset. R (R Core Team, http://www.R-project.org/, Vienna, Austria), Knime (Knime Analytics Platform, https://www.knime.com, Zurich Switzerland), and Rapid Miner (Rapidminer Inc, https://rapidminer.com, Boston, MA) all have packages or nodes that provide different models that can be evaluated for classification problems. By using the training data subset, a predictive model is created and used to classify the revising data after model generation. The preliminary results are evaluated by evaluating how well the revising data were classified. An iterative process follows, allowing for model adjustments, predicting results from the revised data subset, and evaluating results. Model modifications during this process may include decision tree pruning, variable inclusion or exclusion, changing the number of iterations in a Bayesian approach, or other model configuration settings that may affect final classification accuracy. Classification errors (false positives/false negatives) are frequently not equally weighted in terms of the level of concern. This revision process can help ensure classification errors are distributed in a way that is consistent with the overall project objectives. This process can be repeated several times, allowing the model to be optimized for the revised data subset. This framework can be used to evaluate additional classes of classification models. A logistic regression, a single decision tree, and a random forest, for example, could all be considered for their ability to classify a response. Following the tuning process for each type of model, the validation/testing data subset would be used with each model, and the chosen metric (i.e., overall accuracy, sensitivity, specificity) would be compared among the models to determine the best predictive model. Model interpretability, run time, and ease of deployment based on the expected use case are also considered in the final model selection process.

**Validation**

It is not enough to develop a machine learning model and rely on its predictions; you must also check and validate the model's accuracy to ensure the precision of the model's results and make it usable in real-world applications. Choosing the appropriate validation method is also critical to ensuring the accuracy and biases of the validation process. For example, validation may not be required if the data volume is large enough to represent the entire population. However, the situation is different in the real world because the sample or training data sets we are working on may not accurately portray the population. Validation checks for data anomalies, ensure that the data schema hasn't changed, and ensures that the statistics of our new datasets still match those of our previous training datasets.

**Deployment and model monitoring**

Model deployment simply means integrating a machine learning model and integrating it into an existing production environment (1) where it can take in an input and return an output. The goal of deploying your model is to make predictions from a trained ML model available to others, such as users, management, or other systems. Model deployment is closely related to the architecture of ML systems, which refers to the arrangement and interactions of software components within a system to achieve a predefined goal (Opeyemi *et al*., 2019) [79]. Your machine learning model must meet a couple of criteria before it is ready for deployment. 1) Portability: This refers to your software's ability to be transferred from one machine or system to another. A portable model has a low response time and can be rewritten with little effort. 2) Scalability: the ability of your model to scale in size. A scalable model is one that does not require redesigning to maintain its performance. Model monitoring is done after deployment. It is a set of techniques for observing ML models in production and ensuring their performance reliability. Machine learning model monitoring aims to continuously assess the quality of machine learning models in production using data science and statistical techniques. Monitoring can be used for various purposes, such as detecting instabilities early, understanding how and why model performance degrades, and diagnosing specific failure cases.

**Current Challenges in high throughput phenotyping and big data analytics**

Currently, commercial sensors for reliable prediction and disease management in livestock farming via continuous automated real-time monitoring are severely limited. For example, there are no sensors available to measure biomarkers from cows' and pigs' breathing spaces, which could indicate the gut microbiota or even the animals' metabolic states. This gap necessitates the development of sensors and biosensing tools that use "omics" and non-omics approaches to measure biomarkers, miRNAs, and other volatile metabolites, among other things. There are also specific technical challenges, such as where the sensor will be placed, the sampling rate, and how the data will be transmitted. All of these factors influence the accuracy of the algorithms, as well as the scalability and practicability of the solution, which could thus be used on the animal farm. Evaluating sensor position, sampling frequency, sensor data analysis, and window size for data processing would significantly improve farm animal behavior prediction.

**Choosing Machine Learning Algorithms for Data Analysis**

What and how many types of ML features are required, and which algorithms are thus best to tackle the classification problem are decisions that can determine the desired outcome of the animal welfare evaluation. For example, from a set of 44 features, only five to seven may be required to produce highly accurate results. As a result, large feature sets in real-time systems may be problematic due to computational complexity and increased storage requirements. Aside from energy concerns, one of the most significant technical challenges for real-time and long-term farm animal behavior monitoring is "concept drifts." Concept drift occurs when a sensor platform and data analysis system are required to adapt to a change in data distributions within a concept. It is commonly assumed in supervised classification problems that the data in the design model is drawn randomly from the same distribution as the points to be classified in the future. Because of the dynamic nature of many different classification problems, this is an unrealistic assumption. For example, when a system is trained in a single environment, the behavioral classification of animals can show performance discrepancies due to environment variance heterogeneity. Such disparities can be attributed to animal differences (age, breed, etc.) and environmental factors (change in weather conditions, terrain elevation, type of soil, particular farm constraints, etc.).

**Data Sharing**

Defined challenges of big data in livestock include how to share data across institutes and private entities, as well as how to standardize data recording, management, quality control, trait terminology, and data processing. Innovative approaches will be required to distinguish information from noise in data, develop accurate prediction models, and integrate information from disparate sources and locations.

**In Search of More Complex and Better Results**

We are now in the era of sensors, big data, and machine learning. These advanced technologies are expected to drive improved efficiencies and more significant gains in animal farming over the next decade. It will also reduce human errors. As a result, productivity, farmer profits, and animal well-being will improve. More importantly, it has the potential to go beyond improved profits and productivity by assisting us in achieving better animal well-being outcomes. It could also aid in developing more holistic, humane, and environmentally friendly practices.

**Conclusion**

In summary, agricultural HTP technology has the potential to solve the breeder's equation for maximum genetic gain by increasing the intensity and precision of selection, improving genetic variation detection, and shortening breeding cycles. The emergence of Agriculture 4.0 is accelerating the adoption of sensing technologies, big data, and machine learning in modern animal farming. Real-time 24/7 insights into livestock activity, consumption, and production are required in pandemic scenarios where restrictions make it difficult for veterinarians, nutritionists, and producers to visit farms, barns, and feed mills. These findings enabled by sensing technologies generate data that can be accessed remotely, resulting in lower costs and improved performance in responding to consumer demands. Despite the rapid development of AI and machine learning algorithms, there is a global lack of standardization in data collection and sharing. Defined challenges of big data in livestock include sharing data across institutes and private entities, standardizing data recording, management, QC, trait terminology, and data processing. Innovative approaches will be required to distinguish information in data from noise, develop accurate prediction models, and integrate information from disparate sources and locations. Opportunities to use crowdsourcing, machine learning-based artificial intelligence, and other innovative data transfer and storage methods will be critical for extracting knowledge from livestock data. Many data sets, such as those generated in the dairy industry during milk testing are already ready for use. Data utilization, reuse, and generation have enormous potential to improve livestock efficiency, welfare, and societal benefit. Deep phenotyping can benefit society greatly by providing detailed basic physiological knowledge that cannot be measured on humans or model organisms regularly. Training people to work with big data will offer enormous opportunities for academia and private industry to develop knowledge and tools to feed a growing world sustainably.

**Ethics approval and consent to participate**
Not Applicable.

**Consent for publication**
Not Applicable.

**Availability of data and material**
Not Applicable.

**Funding**
Not Applicable.

**Authors' contributions**
AP, MP & HK wrote, reviewed, and approved the manuscript with inputs from all co-authors.

**Declaration of Competing Interest**
There are no conflicts of interest to declare.

**Acknowledgments**
All the authors of the manuscript wish to thank the Director and Joint Director (Research), ICAR-Indian Veterinary Research Institute (IVRI), Izatnagar, an institute under the Indian Council of Agricultural Research (ICAR), Department of Agricultural Research and Education (DARE), Government of India for their support and providing infrastructural facility to review this study.

**References**
1. Abbott D. Applied predictive analytics: principles and techniques for the professional data analyst. John Wiley & Sons; c2014.
2. Abell KM, Theurer ME, Larson RL, White BJ, Hardin DK, Randle RF. Predicting bull behavior events in a multiple-sire pasture with video analysis, accelerometers, and classification algorithms. Computers and Electronics in Agriculture. 2017;136:221-227. https://doi.org/10.1016/j.compag.2017.01.030
3. Adriaens I, Friggens NC, Ouweltjes W, Scott H, Aernouts B, Statham J. Productive life span and resilience rank can be predicted from on-farm first-parity sensor time series but not using a common equation across farms. Journal of Dairy Science. 2020;103(8):7155-7171. https://doi.org/10.3168/jds.2019-17826
4. Agriculture & farming news. Farmer's Weekly; Farmer's Weekly SA; c2016 August 16. https://www.farmersweekly.co.za/
5. Ahmad SF, Panigrahi M, Chhotaray S, Pal D, Parida S, Bhushan B, *et al*. Revelation of genomic breed composition in a crossbred cattle of India with the help of Bovine50K BeadChip. Genomics. 2020;112:1531-1535. https://doi.org/10.1016/j.ygeno.2019.08.025
6. Alsahaf A, Azzopardi G, Ducro B, Hanenberg E, Veerkamp RF, Petkov N. Prediction of slaughter age in pigs and assessment of the predictive value of phenotypic and genetic information using random forest. Journal of Animal Science. 2018;96(12):4935-4943. https://doi.org/10.1093/jas/sky359
7. Alsahaf A, Azzopardi G, Ducro B, Hanenberg E, Veerkamp RF, Petkov N. Estimation of muscle scores of live pigs using a Kinect camera. IEEE Access: Practical Innovations, Open Solutions. 2019;7:52238-52245. https://doi.org/10.1109/access.2019.2910986
8. Amrine DE, White BJ, Larson RL. Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for bovine respiratory disease. Computers and Electronics in Agriculture. 2014;105:9-19. https://doi.org/10.1016/j.compag.2014.04.009
9. Bar D, Solomon R. Rumination collars: What can they tell us. Proceedings of the First North American Conference on Precision Dairy Management. 2010;2:214-215.
10. Barkema HW, von Keyserlingk MAG, Kastelic JP, Lam TJGM, Luby C, Roy J-P, *et al*. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. Journal of Dairy Science. 2015;98(11):7426-7445. https://doi.org/10.3168/jds.2015-9377
11. Bicalho RC, Cheong SH, Cramer G, Guard CL. Association between a visual and an automated locomotion score in lactating Holstein cows. Journal of Dairy Science. 2007;90(7):3294-3300. https://doi.org/10.3168/jds.2007-0076
12. Borgonovo F, Ferrante V, Grilli G, Pascuzzo R, Vantini S, Guarino M. A data-driven prediction method for an early warning of coccidiosis in intensive livestock systems: A preliminary study. Animals: An Open Access Journal from MDPI. 2020;10(4):747. https://doi.org/10.3390/ani10040747
13. Brito LF, Oliveira HR, McConn BR, Schinckel AP, Arrazola A, Marchant-Forde JN, *et al*. Large-scale phenotyping of livestock welfare in commercial production systems: A new frontier in animal breeding. Frontiers in Genetics. 2020;11:793. https://doi.org/10.3389/fgene.2020.00793
14. Büchel S, Sundrum A. Short communication: Decrease in rumination time as an indicator of the onset of calving. Journal of Dairy Science. 2014;97(5):3120-3127. https://doi.org/10.3168/jds.2013-7613
15. Bucher S, Sundrum A. Short communication: decrease in rumination time as an indicator of the onset of calving. J Dairy Sci. 2014;97:3120-3127.
16. Burke CR, Meir S, Pellow R. Use of progesterone profiling to investigate factors affecting conception rates in a large, pasture-grazed dairy herd. Proceedings of the New Zealand Society of Animal Production. 2012;72:45-50.
17. Carabús A, Gispert M, Font-i-Furnols M. Imaging technologies to study the composition of live pigs: A review. Revista de Investigacion Agraria [Spanish Journal of Agricultural Research]. 2016;14(3):e06R01. https://doi.org/10.5424/sjar/2016143-8439
18. Chhotaray S, Panigrahi M, Pal D, Ahmad SF, Bhushan B, Gaur GK, *et al*. Ancestry informative markers derived from discriminant analysis of principal components provide important insights into the composition of crossbred cattle. Genomics. 2020;112:1726-1733. https://doi.org/10.1016/j.ygeno.2019.10.008
19. Chhotaray S, Panigrahi M, Bhushan B, Gaur GK, Dutt T, Mishra BP, *et al*. Genome-wide association study reveals genes crucial for coat color production in Vrindavani cattle. Livest. Sci. 2021a;247:104476. https://doi.org/10.1016/j.livsci.2021.104476
20. Chhotaray S, Panigrahi M, Pal D, Ahmad SF, Bhanuprakash V, Kumar H, *et al*. Genome-wide estimation of inbreeding coefficient, effective population size and haplotype blocks in Vrindavani crossbred cattle

strain of India. Biol. Rhythm Res. 2021b;52(5):666-79. https://doi.org/10.1080/09291016.2019.1600266
21. Clark CEF, Lyons NA, Millapan L, Talukder S, Cronin GM, Kerrisk KL, *et al*. Rumination and activity levels as predictors of calving for dairy cows. Animal: An International Journal of Animal Bioscience. 2015;9(4):691-695. https://doi.org/10.1017/S1751731114003127
22. Colditz IG, Hine BC. Resilience in farm animals: biology, management, breeding and implications for animal welfare. Animal Production Science. 2016;56(12):1961. https://doi.org/10.1071/an15297
23. Crossa J, Fritsche-Neto R, Montesinos-Lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-Lopez A, *et al*. The modern plant breeding triangle: Optimizing the use of genomics, phenomics, and enviromics data. Frontiers in Plant Science. 2021;12:651480. https://doi.org/10.3389/fpls.2021.651480
24. De Marchi M, Toffanin V, Cassandro M, Penasa M. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. Journal of Dairy Science. 2014;97(3):1171-1186. https://doi.org/10.3168/jds.2013-6799
25. Dehareng F, Delfosse C, Froidmont E, Soyeurt H, Martin C, Gengler N, *et al*. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. Animal: An International Journal of Animal Bioscience. 2012;6(10):1694-1701. https://doi.org/10.1017/S1751731112000456
26. Dhoble AS, Ryan KT, Lahiri P, Chen M, Pang X, Cardoso FC, *et al*. Cytometric fingerprinting and machine learning (CFML): A novel label-free, objective method for routine mastitis screening. Computers and Electronics in Agriculture. 2019;162:505-513. https://doi.org/10.1016/j.compag.2019.04.029
27. Digital Object Identifier System. (n.d.). Doi.org. Retrieved August 21, 2022, from https://doi.org/
28. Ebrahimi M, Mohammadi-Dehcheshmeh M, Ebrahimie E, Petrovski KR. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. Computers in Biology and Medicine. 2019;114(103456):103456. https://doi.org/10.1016/j.compbiomed.2019.103456
29. Ellis JL, Jacobs M, Dijkstra J, van Laar H, Cant JP, Tulpan D, *et al*. Review: Synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data. Animal: An International Journal of Animal Bioscience. 2020;14(S2):s223-s237. https://doi.org/10.1017/S1751731120000312
30. Eradus WJ, Rossing W, Hogewerf PH, Benders E. Signal processing of activity data for oestrus detection in dairy cattle. In Proceedings of the International Symposium on Prospects for Automatic Milking. Pudoc Scientific; c1992.
31. Ferguson NS. Optimization: A paradigm change in nutrition and economic solutions. Advances in Pork Production. 2014;25:121-127.
32. Fernández-Carrión E, Martínez-Avilés M, Ivorra B, Martínez-López B, Ramos ÁM, Sánchez-Vizcaíno JM. Motion-based video monitoring for early detection of livestock diseases: The case of African swine fever. PloS One. 2017;12(9):e0183793. https://doi.org/10.1371/journal.pone.0183793
33. Finkel E. With ‘phenomics,’plant scientists hope to shift breeding into overdrive; c2009.
34. Gengler N, Soyeurt H, Dehareng F, Bastin C, Colinet F, Hammami H, *et al*. Capitalizing on fine milk composition for breeding and management of dairy cows. Journal of Dairy Science. 2016;99(5):4071-4079. https://doi.org/10.3168/jds.2015-10140
35. Gertz M, Große-Butenuth K, Junge W, Maassen-Francke B, Renner C, Sparenberg H, *et al*. Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. Computers and Electronics in Agriculture. 2020;173(105404):105404. https://doi.org/10.1016/j.compag.2020.105404
36. Gillespie J (n.d.). Milk cost of production estimates. Usda.gov. Retrieved August 19, 2022, from http://www.ers.usda.gov/data-products/milk-cost-of-production-estimates.aspx
37. Green ED, Watson JD, Collins FS. Human Genome Project: Twenty-five years of big biology. Nature. 2015;526(7571):29-31. https://doi.org/10.1038/526029a
38. Green LE, Hedges VJ, Schukken YH, Blowey RW, Packington AJ. The impact of clinical lameness on the milk yield of dairy cows. Journal of Dairy Science. 2002;85(9):2250-2256. https://doi.org/10.3168/jds.S0022-0302(02)74304-X
39. Grelet C, Bastin C, Gelé M, Davière J-B, Johan M, Werner A, *et al*. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. Journal of Dairy Science. 2016;99(6):4816-4825. https://doi.org/10.3168/jds.2015-10477
40. Grzesiak W, Błaszczyk P, Lacroix R. Methods of predicting milk yield in dairy cows: predictive capabilities of Wood’s lactation curve and artificial neural networks (ANNs). Comput. Electron. Agri. 2006;54:69-83. https://doi.org/10.1016/j.compag.2006.08.004
41. Guarino M, Norton T, Berckmans D, Vranken E, Berckmans D. A blueprint for developing and applying precision livestock farming tools: A key output of the EU-PLF project. Animal Frontiers. 2017;7(1):12-17. https://doi.org/10.2527/af.2017.0103
42. Hartwell L, Mankoff D, Paulovich A, Ramsey S, Swisher E. Cancer biomarkers: a systems approach. Nature Biotechnology. 2006;24(8):905-908. https://doi.org/10.1038/nbt0806-905
43. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. PDF) High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. J Dairy Sci. 2009;92(2):433-443.
44. Hayes T, Usami S, Jacobucci R, McArdle JJ. Using Classification and Regression Trees (CART) and random forests to analyze attrition: Results from two simulations. Psychology and Aging. 2015;30(4):911-929. https://doi.org/10.1037/pag0000046
45. Heffner EL, Lorenz AJ, Jannink J, Sorrells ME. Plant breeding with genomic selection: Gain per unit time and cost. Crop Science. 2010;50(5):1681-1690. https://doi.org/10.2135/cropsci2009.11.0662
46. Hein NT, Ciampitti IA, Jagadish SVK. Bottlenecks and

opportunities in field-based high-throughput phenotyping for heat and drought stress. Journal of Experimental Botany. 2021;72(14):5102-5116. https://doi.org/10.1093/jxb/erab021
47. Hempstalk K, Harris BL, Lopdell TJ. Automatic reading of Kamars for heat detection in dairy herds. Proceedings of the New Zealand Society of Animal Production. 2010;70:8-12.
48. Hidalgo A, Zouari F, Knijn H, Van Der Beek S. Prediction of postpartum diseases of dairy cattle using machine learning. In Proceedings of the World Congress on Genetics Applied to Livestock Production. World Congress on Genetics Applied to Livestock Production; c2018.
49. Hill WG. Is continued genetic improvement of livestock sustainable? Genetics. 2016;202(3):877-881. https://doi.org/10.1534/genetics.115.186650
50. Ida N, Icdla P, Mds B, Agd M, Kaod L, Lgdf B, *et al*. Assessing locomotion deficiency in broiler chicken. Scientia Agricola. 2010;67:129-135.
51. Juarez ST, Robinson PH, DePeters EJ, Price EO. Impact of lameness on behavior and productivity of lactating Holstein cows. Applied Animal Behaviour Science. 2003;83(1):1-14. https://doi.org/10.1016/s0168-1591(03)00107-2
52. Kaisa K, Kumar H, Saravanan K, Rajawat D, Bhushan B, Kumar P, *et al*. Concepts of Genomic Selection in Poultry and its Applications. Int. J Livest. Res. 2020;10(10):32-42. https://doi.org/10.5455/ijlr.20200427022022
53. Kamphuis C, Duenk P, Veerkamp RF, Visser B, Singh G, Nigsch A, *et al*. Machine learning to further improve the decision which boar ejaculates to process into artificial insemination doses. Theriogenology. 2020;144:112-121. https://doi.org/10.1016/j.theriogenology.2019.12.017
54. Kellenberger B, Marcos D, Tuia D. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. Remote Sensing of Environment. 2018;216:139-153. https://doi.org/10.1016/j.rse.2018.06.028
55. Knowlton KF, Radcliffe JS, Novak CL, Emmerson DA. Animal management to reduce phosphorus losses to the environment. Journal of Animal Science. 2004;82E(13):E173-195. https://doi.org/10.2527/2004.8213_supplE173x
56. Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA. Challenges in combining projections from multiple climate models. Journal of Climate. 2010;23(10):2739-2758. https://doi.org/10.1175/2009jcli3361.1
57. Kongsro J. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Computers and Electronics in Agriculture. 2014;109:32-35. https://doi.org/10.1016/j.compag.2014.08.008
58. König S, May K. Invited review: Phenotyping strategies and quantitative-genetic background of resistance, tolerance and resilience associated traits in dairy cattle. Animal: An International Journal of Animal Bioscience. 2019;13(5):897-908. https://doi.org/10.1017/S1751731118003208
59. Kumar H, Panigrahi M, Chhotaray S, Pal D, Bhanuprakash V, Saravanan KA, *et al*. Identification of breed-specific SNP panel in nine different cattle genomes. Biomed. Res. 2019;30(1). https://doi.org/10.35841/biomedicalresearch.30-18-1195
60. Kumar H, Panigrahi M, Chhotaray S, Parida S, Chauhan A, Bhushan B, *et al*. Comparative analysis of five different methods to design a breed-specific SNP panel for cattle. Anim. Biotechnol. 2021a;32(1):130-6. https://doi.org/10.1080/10495398.2019.1646266
61. Kumar H, Panigrahi M, Panwar A, Rajawat D, Nayak SS, Saravanan KA, *et al*. Machine-Learning Prospects for Detecting Selection Signatures Using Population Genomics Data. J Comput. Biol; c2022. https://doi.org/10.1089/cmb.2021.0447
62. Kumar H, Panigrahi M, Rajawat D, Panwar A, Nayak SS, Kaisa K, *et al*. Selection of breed-specific SNPs in three Indian sheep breeds using ovine 50 K array. Small Rumin. Res. 2021b;205:106545. https://doi.org/10.1016/j.smallrumres.2021.106545
63. Kumar H, Panigrahi M, Saravanan KA, Rajawat D, Parida S, Bhushan B, *et al*. Genome-wide detection of copy number variations in Tharparkar cattle. Anim. Biotechnol; c2021c. p. 1-8. https://doi.org/10.1080/10495398.2021.1942027
64. Kumar H, Panigrahi M, Saravanan KA, Parida S, Bhushan B, Gaur GK, *et al*. SNPs with intermediate minor allele frequencies facilitate accurate breed assignment of Indian Tharparkar cattle. Gene. 2021d;777:145473. https://doi.org/10.1016/j.gene.2021.145473
65. Larson RL, White BJ. First steps when using scientific literature in clinical veterinary practice. J Am. Vet. Med. Assoc. 2015;247:254-258. https://doi.org/10.2460/javma.247.3.254
66. Lee J, Jin L, Park D, Chung Y. Automatic recognition of aggressive behavior in pigs using a Kinect depth sensor. Sensors (Basel, Switzerland). 2016;16(5). https://doi.org/10.3390/s16050631
67. Matthews SG, Miller AL, Clapp J, Plötz T, Kyriazakis I. Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. Veterinary Journal (London, England: 1997). 2016;217:43-51. https://doi.org/10.1016/j.tvjl.2016.09.005
68. Mcgowan JE, Burke CR, Jago JG. Validation of a technology for objectively measuring behaviour in dairy cows and its application for oestrous detection. Proceedings of the New Zealand Society of Animal Production. 2007;67:136-142.
69. Mehrotra A, Bhushan B, Karthikeyan A, Singh A, Panda S, Bhati M, *et al*. Genome-wide SNP data unravel the ancestry and signatures of divergent selection in Ghurrah pigs of India. Livest Sci. 2021a;250:104587. https://doi.org/10.1016/j.livsci.2021.104587.
70. Mehrotra A, Bhushan B, Kumar A, Panigrahi MAK, Singh A, Tiwari AK, *et al*. A 1.6 Mb region on SSC2 is associated with antibody response to classical swine fever vaccination in a mixed pig population. Anim. Biotechnol; c2021b. p. 1-6. Advance online publication. https://doi.org/10.1080/10495398.2021.1873145
71. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819
72. Miedema HM, Cockram MS, Dwyer CM, Macrae AI. Changes in the behaviour of dairy cows during the 24 h before normal calving compared with behaviour during late pregnancy. Appl. Anim. Behav. Sci. 2011;131:8-14.

73. Mir RR, Choudhary N, Singh B, Khandy IA, Bawa V, Sofi P, *et al*. Harnessing Genomics Through Phenomics. In Phenomics in Crop Plants: Trends, Options and Limitations. Springer India; c2015. p. 273-283.
74. Mollenhorst H, Ducro BJ, De Greef KH, Hulsegge I, Kamphuis C. Boosted trees to predict pneumonia, growth, and meat percentage of growing-finishing pigs1. Journal of Animal Science. 2019;97(10):4152-4159. https://doi.org/10.1093/jas/skz274
75. Morota G, Ventura RV, Silva FF, Koyama M, Fernando SC. BIG data analytics and precision animal agriculture symposium: machine learning and data mining advance predictive big data analysis in precision animal agriculture. J Anim. Sci. 2018;96:1540-1550.
76. Mulder HA, Rashidi H. Selection on resilience improves disease resistance and tolerance to infections. Journal of Animal Science. 2017;95(8):3346-3358. https://doi.org/10.2527/jas.2017.1479
77. Nääs I de A, Paz IC de LA, Baracho M dos S, Menezes AG de, Lima KAO de, Bueno LG de F, *et al*. Assessing locomotion deficiency in broiler chicken. Scientia Agricola. 2010;67(2):129-135. https://doi.org/10.1590/s0103-90162010000200001
78. Neethirajan S. The role of sensors, big data and machine learning in modern animal farming. Sensing and Bio-Sensing Research. 2020;29(100367):100367. https://doi.org/10.1016/j.sbsr.2020.100367
79. Opeyemi B. Deployment of Machine learning Models Demystified (Part 1). Towards Data Science; c2019.
80. Ouellet V, Vasseur E, Heuwieser W, Burfeind O, Maldague X, Charbonneau É. Evaluation of calving indicators measured by automated monitoring devices to predict the onset of calving in Holstein dairy cows. Journal of Dairy Science. 2016;99(2):1539-1548. https://doi.org/10.3168/jds.2015-10057
81. Ouweltjes W, Spoelstra M, Ducro B, de Haas Y, Kamphuis C. A data-driven prediction of lifetime resilience of dairy cows using commercial sensor data collected during first lactation. Journal of Dairy Science. 2021;104(11):11759-11769. https://doi.org/10.3168/jds.2021-20413
82. Pal D, Panigrahi M, Chhotaray S, Kumar H, Nayak SS, Rajawat D, *et al*. Unraveling genetic admixture in the Indian crossbred cattle by different approaches using Bovine 50K BeadChip. Trop. Anim. Health Prod. 2022;54(2):135. https://doi.org/10.1007/s11250-022-03133-7
83. Panigrahi M, Kumar H, Nayak SS, Rajawat D, Parida S, Bhushan B, *et al*. Molecular characterization of CRBR2 fragment of TLR4 gene in association with mastitis in Vrindavani cattle. Microb. Pathog. 2022a;165:105483. https://doi.org/10.1016/j.micpath.2022.105483.
84. Panigrahi M, Kumar H, Sah V, Dillipkumar Verma A, Bhushan B, Parida S. Transcriptome profiling of buffalo endometrium reveals molecular signature distinct to early pregnancy. Gene. 2020;743:144614. https://doi.org/10.1016/j.gene.2020.144614
85. Panigrahi M, Kumar H, Saravanan KA, Rajawat D, Sonejita Nayak S, Ghildiyal K, *et al*. Trajectory of livestock genomics in South Asia: A comprehensive review. Gene. 2022b;843:146808. Advance online publication. https://doi.org/10.1016/j.gene.2022.146808
86. Parasar P, Bhushan B, Panigrahi M, Kumar H, Kaisa K, Dutt T. Characterization of BoLA class II DQA and DQB by PCR-RFLP, cloning, and sequencing reveals sequence diversity in crossbred cattle. Anim. Biotechnol; c2021. p. 1-11. Advance online publication. https://doi.org/10.1080/10495398.2021.2006205
87. Patra B, Panigrahi M, Kumar H, Kaisa K, Dutt T, Bhushan B. Molecular and phylogenetic analysis of MHC class I exons 7-8 in a variety of cattle and buffalo breeds. Anim. Biotechnol; c2021. p. 1-7. Advance online publication. https://doi.org/10.1080/10495398.2021.1999969
88. Pezzuolo A, Guarino M, Sartori L, González LA, Marinello F. On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera. Computers and Electronics in Agriculture. 2018;148:29-36. https://doi.org/10.1016/j.compag.2018.03.003
89. Piñeiro JM, Menichetti BT, Barragan AA, Relling AE, Weiss WP, Bas S, *et al*. Associations of pre-and postpartum lying time with metabolic, inflammation, and health status of lactating dairy cows. Journal of Dairy Science. 2019;102(4):3348-3361.
90. Pomar C, Remus A. Precision pig feeding: a breakthrough toward sustainability. Animal Frontiers. 2019;9(2):52-59. https://doi.org/10.1093/af/vfz006
91. Poppe M, Veerkamp RF, van Pelt ML, Mulder HA. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. Journal of Dairy Science. 2020;103(2):1667-1684. https://doi.org/10.3168/jds.2019-17290
92. Rahaman MM, Chen D, Gillani Z, Klukas C, Chen M. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. Frontiers in Plant Science. 2015;6:619. https://doi.org/10.3389/fpls.2015.00619
93. Rajawat D, Panigrahi M, Kumar H, Nayak SS, Parida S, Bhushan B, *et al*. Identification of important genomic footprints using eight different selection signature statistics in domestic cattle breeds. Gene. 2022a;816:146165. https://doi.org/10.1016/j.gene.2021.146165
94. Rajawat D, Panigrahi M, Kumar H, Nayak SS, Saravanan KA, Bhushan B, *et al*. Revealing Genomic Footprints of Selection for Fiber and Production Traits in Three Indian Sheep Breeds. J Nat. Fibers; c2022b. Doi: 10.1080/15440478.2022.2069198
95. Rojas-Downing MM, Nejadhashemi AP, Harrigan T, Woznicki SA. Climate change and livestock: Impacts, adaptation, and mitigation. Climate Risk Management. 2017;16:145-163. https://doi.org/10.1016/j.crm.2017.02.001
96. Rutten CJ, Kamphuis C, Hogeveen H, Huijps K, Nielen M, Steeneveld W. Sensor data on cow activity, rumination, and ear temperature improve prediction of the start of calving in dairy cows. Computers and Electronics in Agriculture. 2017;132:108-118. https://doi.org/10.1016/j.compag.2016.11.009
97. Saint-Dizier M, Chastant-Maillard S. Methods and on-farm devices to predict calving time in cattle. Veterinary Journal (London, England. 2015, 1997;205(3):349-356. https://doi.org/10.1016/j.tvjl.2015.05.006
98. Saravanan KA, Panigrahi M, Kumar H, Rajawat D, Nayak SS, Bhushan B, *et al*. Role of genomics in combating COVID-19 pandemic. Gene. 2022b;823:146387.

https://doi.org/10.1016/j.gene.2022.146387
99. Saravanan KA, Panigrahi M, Kumar H, Bhushan B. Advanced software programs for the analysis of genetic diversity in livestock genomics: a mini review. Biol. Rhythm Res; c2019. p. 1-11. https://doi.org/10.1080/09291016.2019.1642650.
100. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. Genome-wide analysis of genetic diversity and selection signatures in three Indian sheep breeds. Livest. Sci. 2021a;243:104367. https://doi.org/10.1016/j.livsci.2020.104367
101. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. Selection signatures in livestock genome: A review of concepts, approaches and applications. Livest. Sci. 2020a;241:104257. https://doi.org/10.1016/j.livsci.2020.104257
102. Saravanan KA, Panigrahi M, Kumar H, Nayak SS, Rajawat D, Bhushan B, *et al*. Progress and future perspectives of livestock genomics in India: a mini review. Anim. Biotechnol; c2022a. https://doi.org/10.1080/10495398.2022.2056046
103. Saravanan KA, Panigrahi M, Kumar H, Parida S, Bhushan B, Gaur GK, *et al*. Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. Anim. Biotechnol; c2020b. p. 1-15. https://doi.org/10.1080/10495398.2020.1796696
104. Saravanan KA, Panigrahi M, Kumar H, Parida S, Bhushan B, Gaur GK, *et al*. Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. Genomics; 2021b;113:955-963. https://doi.org/10.1016/j.ygeno.2021.02.009
105. Satopää VA, Baron J, Foster DP, Mellers BA, Tetlock PE, Ungar LH. Combining multiple probability predictions using a simple logit model. International Journal of Forecasting. 2014;30(2):344-356. https://doi.org/10.1016/j.ijforecast.2013.09.009
106. Schirmann K, Chapinal N, Weary DM, Vickers L, von Keyserlingk MAG. Short communication: Rumination and feeding behavior before and after calving in dairy cows. Journal of Dairy Science. 2013;96(11):7088-7092. https://doi.org/10.3168/jds.2013-7023
107. Scholz AM, Bünger L, Kongsro J, Baulain U, Mitchell AD. Non-invasive methods for the determination of body and carcass composition in livestock: dual-energy X-ray absorptiometry, computed tomography, magnetic resonance imaging and ultrasound: invited review. Animal: An International Journal of Animal Bioscience. 2015;9(7):1250-1264. https://doi.org/10.1017/s1751731115000336
108. Sinha A, Chen H, Danu DG, Kirubarajan T, Farooq M. Estimation and decision fusion: A survey. Neurocomputing. 2008;71(13-15):2650-2656. https://doi.org/10.1016/j.neucom.2007.06.016
109. Sprecher DJ, Hostetler DE, Kaneene JB. A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance. Theriogenology. 1997;47(6):1179-1187. https://doi.org/10.1016/s0093-691x(97)00098-8
110. Stavrakakis S, Li W, Guy JH, Morgan G, Ushaw G, Johnson GR, *et al*. Validity of the Microsoft Kinect sensor for assessment of normal walking patterns in pigs. Computers and Electronics in Agriculture. 2015;117:1-7. https://doi.org/10.1016/j.compag.2015.07.003
111. Taneja M, Byabazaire J, Jalodia N, Davy A, Olariu C, Malone P. Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle. Computers and Electronics in Agriculture. 2020;171(105286):105286. https://doi.org/10.1016/j.compag.2020.105286
112. Tranter WP, Morris RS. A case study of lameness in three dairy herds. New Zealand Veterinary Journal. 1991;39(3):88-96. https://doi.org/10.1080/00480169.1991.35668
113. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. Nature Reviews. Drug Discovery. 2007;6(4):287-293. https://doi.org/10.1038/nrd2251
114. Tucker CB, Jensen MB, de Passillé AM, Hänninen L, Rushen J. Invited review: Lying time and the welfare of dairy cows. Journal of Dairy Science. 2021;104(1):20-46. https://doi.org/10.3168/jds.2019-18074
115. van der Heide EMM, Kamphuis C, Veerkamp RF, Athanasiadis IN, Azzopardi G, van Pelt ML, *et al*. Improving predictive performance on survival in dairy cattle using an ensemble learning approach. Computers and Electronics in Agriculture. 2020;177(105675):105675. https://doi.org/10.1016/j.compag.2020.105675
116. VanderWaal K, Morrison RB, Neuhauser C, Vilalta C, Perez AM. Translating big data into smart data for veterinary epidemiology. Frontiers in Veterinary Science. 2017;4:110. https://doi.org/10.3389/fvets.2017.00110
117. von Keyserlingk MAG, Martin NP, Kebreab E, Knowlton KF, Grant RJ, Stephenson M, *et al*. Invited review: Sustainability of the US dairy industry. Journal of Dairy Science. 2013;96(9):5405-5425. https://doi.org/10.3168/jds.2012-6354
118. Warner D, Vasseur E, Lefebvre DM, Lacroix R. A machine learning based decision aid for lameness in dairy herds using farm-based records. Computers and Electronics in Agriculture. 2020;169(105193):105193. https://doi.org/10.1016/j.compag.2019.105193
119. Weersink A, Fraser E, Pannell D, Duncan E, Rotz S. Opportunities and challenges for Big Data in agricultural and environmental analysis. Annual Review of Resource Economics. 2018;10(1):19-37. https://doi.org/10.1146/annurev-resource-100516-053654
120. Whay HR, Waterman AE, Webster AJ. Associations between locomotion, claw lesions and nociceptive threshold in dairy heifers during the peri-partum period. Veterinary Journal (London, England: 1997). 1997;154(2):155-161. https://doi.org/10.1016/s1090-0233(97)80053-6
121. Wolfert S, Ge L, Verdouw C, Bogaardt MJ. Big data in smart farming-a review. Agricultural Systems. 2017a;153:69-80.
122. Wolfert S, Ge L, Verdouw C, Bogaardt M-J. Big data in smart farming- a review. Agric. Syst. 2017b;153:69-80.
123. Woźniak M, Graña M, Corchado E. A survey of multiple classifier systems as hybrid systems. An International Journal on Information Fusion. 2014;16:3-17. https://doi.org/10.1016/j.inffus.2013.04.006
124. Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, *et al*. Crop phenomics and highthroughput phenotyping: past decades, current

challenges, and future perspectives. Molecular Plant. 2020;13:187-214.
125. Yoo W, Mayberry R, Bae S, Singh K, Peter He Q, Lillard JW Jr. A study of effects of MultiCollinearity in the multivariable analysis. International Journal of Applied Science and Technology. 2014;4(5):9-19.
126. Ziegler A, Koch A, Krockenberger K, Grosshennig A. Personalized medicine using DNA biomarkers: a review. Human Genetics. 2012;131(10):1627-1638. https://doi.org/10.1007/s00439-012-1188-9
127. Titler M, Maquivar MG, Bas S, Rajala-Schultz PJ, Gordon E, McCullough K, *et al*. Prediction of parturition in Holstein dairy cattle using electronic data loggers. Journal of Dairy Science. 2015 Aug 1;98(8):5304-12.
128. Smith VH, Joye SB, Howarth RW. Eutrophication of freshwater and marine ecosystems. Limnology and oceanography. 2006 Jan;51(1-2):351-5.
129. Perkmann T, Mucher P, Perkmann-Nagele N, Radakovics A, Repl M, *et al*. The comparability of Anti-Spike SARS-CoV-2 antibody tests is time-dependent: a prospective observational study. Microbiology Spectrum. 2022 Feb 23;10(1):e01402-21.
130. Probert AF, Wegmann D, Volery L, Adriaens T, Bakiu R, Bertolino S, *et al*. Identifying, reducing, and communicating uncertainty in community science: a focus on alien species. Biological Invasions. 2022 Aug 29:1-27.
131. Yan Y, Gause KT, Kamphuis MM, Ang CS, O'Brien-Simpson NM, Lenzo JC, *et al*. Differential roles of the protein corona in the cellular uptake of nanoporous polymer particles by monocyte and macrophage cell lines. ACS nano. 2013 Dec 23;7(12):10960-70.
132. Flores-González JC, Mayordomo-Colunga J, Jordan I, Miras-Veiga A, Montero-Valladares C, Olmedilla-Jodar M, *et al*. Prospective multicentre study on the epidemiology and current therapeutic management of severe bronchiolitis in Spain. BioMed research international. 2017 Mar 22;2017.