

## Article

# Hierarchical Clustering on Principal Components Analysis to Detect Clusters of Highly Pathogenic Avian Influenza Subtype H5N6 Epidemic across South Korean Poultry Farms

Kye-Young Koh <sup>1,†</sup> , Saleem Ahmad <sup>1,†</sup> , Jae-il Lee <sup>1</sup> , Guk-Hyun Suh <sup>2</sup> and Chang-Min Lee <sup>2,\*</sup>

<sup>1</sup> Veterinary Public Health Lab, College of Veterinary Medicine, Chonnam National University, Gwangju 61186, Korea; 178510@jnu.ac.kr (K.-Y.K.); 186762@jnu.ac.kr (S.A.); jaeil@jnu.ac.kr (J.-i.L.)

<sup>2</sup> Department of Veterinary Internal Medicine, College of Veterinary Medicine and BK21 FOUR Program, Chonnam National University, Gwangju 61186, Korea; ghsuh@chonnam.ac.kr

\* Correspondence: cmlee1122@jnu.ac.kr

† These authors contributed equally to this work.

**Abstract:** Several outbreaks of highly pathogenic avian influenza (HPAI) in poultry have already been documented across the world, causing major economic losses. Research on diverse perspectives for future HPAI outbreaks' prevention is desperately needed. It is critical to determine high-risk areas for HPAI outbreaks in order to develop high-level biosecurity in all such areas. The aim of this study is to identify high-risk areas as hotspots for high rates of birds' infection and mortality and culling. We used "hierarchical clustering on principal components" (HCPC) to classify infected poultry farms in South Korea based on the point prevalence rate, infections, and deaths in susceptible birds. The linear combination of the original predictors was determined using "principal component analysis (PCA)". Based on PCA, we applied the hierarchical clustering algorithm, which divided the data into four clusters based on the dissimilarity matrix. These four groups of poultry farms were identified on the basis of five variables. According to the findings based on the HCPC method, poultry farms in "cluster 4" had significantly higher average bird infections with high mortality when compared to other clusters. Similarly, the poultry farms in "cluster 2" had robust average bird culling in place to limit bird infectivity and mortality due to a high number of susceptible birds. The poultry farms belonging to "cluster 3" had a significantly higher average point prevalence rate of HPAI H5N6 cases than the rest of the clusters. Based on this study, it is recommended that poultry farms with a high number of infections and mortality in susceptible birds should implement proper biosecurity management to control HPAI infections while avoiding the culling of a large number of birds.

**Keywords:** highly pathogenic avian influenza; epidemic; hierarchical clustering on principal components; prevalence; biosecurity management; prevention



**Citation:** Koh, K.-Y.; Ahmad, S.; Lee, J.-i.; Suh, G.-H.; Lee, C.-M.

Hierarchical Clustering on Principal Components Analysis to Detect Clusters of Highly Pathogenic Avian Influenza Subtype H5N6 Epidemic across South Korean Poultry Farms. *Symmetry* **2022**, *14*, 598. <https://doi.org/10.3390/sym14030598>

Academic Editor: Stefan Van Dongen

Received: 14 January 2022

Accepted: 14 March 2022

Published: 17 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Influenza virus subtypes produce avian influenza, which is an infectious disease that affects birds. The avian influenza virus (AIV) was discovered almost a century ago. A variety of avian influenza viruses infect domesticated and wild birds throughout the world. Virus types that are deadly to birds have evolved significantly over the years and are known to spread rapidly through flocks and cause significant mortality in poultry [1].

Influenza A virus (IAV) is a multi-host virus that has serious consequences for wild and domesticated animals as well as humans [2]. Ducks, geese, gulls and terns, in particular, appear crucial in the propagation of the influenza A virus [3,4]. AIV may be spread from wild birds to domestic birds, causing significant economic disruption, and it can even be transmitted to humans in rare cases, rendering it a zoonotic disease [5].

According to Chatziprodromidou et al. [6], Asia is the most common continent for AIV outbreaks. During the period of 2010–2016, various Asian countries, including South Korea, recorded some of the deadliest HPAI outbreaks.

IAVs not only cause economic implications in poultry but also influence the evolution of advanced human strains to some extent through genetic reassortment; it is an essential factor to consider in terms of public health [7,8]. The global poultry market will be severely impacted by contemporaneous outbreaks of HPAI in Asia as well as Europe [9].

Various AIV H5 subtypes were found more prominent in wild bird cases and are widespread due to their migration [10].

Avian influenza is an A-type influenza virus, belonging to the Orthomyxoviridae family of viruses. Because of the interaction of the primary virion antigens, haemagglutinin (HA) and neuraminidase (NA), type-A influenza has hundreds of distinct antigen variants [11]. To date, 16 HA subtypes and 9 NA subtypes are recognized, with two more HA and NA subtypes reported in bats [12,13]. The term “highly pathogenic avian influenza” (HPAI) usually applies to variants that can cause an “intravenous pathogenicity index” (IVPI) more than 1.2 or a death rate higher than 75% in a specific poultry population during a 10-day period. According to this criterion, all HPAI strains identified so far are of either the H5 or H7 subtypes [6].

Avian influenza viruses are transmitted in bird feces and respiratory secretions, although the volume of virus secreted varies depending on the virus, host species and other variables [14–16].

Domesticable birds, including chickens and turkeys, can contract AIVs through close interaction with affected waterfowl or poultry. AIVs infect around 105 bird species worldwide; however, the natural reservoirs for this virus are water birds, including shorebirds, terns and gulls [17].

The highly pathogenic avian influenza (HPAI) virus (HPAIV), a subtype of influenza virus A, is a virus that infects poultry, including wild birds. The virus, as the name implies, produces high fatality rates in susceptible birds. Distinctive, genetically diverse viruses are continually emerging due to the viral genome’s eight RNA segments, posing a challenge to domestic chicken businesses globally [18].

H5N6 HPAIVs are constantly developing in South Korea and thus are detected in wild migratory birds [19]. Apparently, H5N6 HPAIVs are linked to human diseases [20]. Since 2003, all HPAIVs detected have been H5-subtype viruses, which are genetically distinct from each other as well, and the majority of them have been imported from outside Korea. Culling is the procedure of eliminating diseased and unaffected poultry in regions surrounding an affected zone in order to quickly restrict the spread of an infectious disease outbreak, with the goal of eventually eradicating the malady [21]. In fact, authorities mainly apply the culling techniques outlined below, which are based on zoning rules.

Culling on infected premises (IP): eliminating all poultry on a poultry farm or in a local region where affected or dead birds have been identified as avian-influenza-positive.

Preemptive (PE) culling: the practice of eliminating all poultry birds within a defined area, which, in the Republic of Korea, is designated as a 3 km radius around the IP.

Even though these strategies have been regarded as one of the principal control methods for years, there have been some concerns regarding the welfare of animals and profitability. Particularly when PE culling is carried out in densely packed farm areas, a great percentage of chickens are ultimately eliminated. In 2008, the Republic of Korea eliminated around 10 million poultry to tackle the H5N1 epidemic [22].

The ongoing HPAI epidemics in Korea caused by novel H5Nx viruses are intimately related to migratory birds, and the majority of HPAIV strains are genetically identical to those found in Southeast Asian countries, including China [23]. As AI viruses continue to evolve and reassort in China’s live poultry markets, the development of various AI viruses in China is inevitable [24]; potential HPAI outbreaks caused by significant new HPAIVs may be unavoidable in Korea.

The newly isolated H5N6 virus clade 2.3.4.4.B is phylogenetically different from the previously discovered clade 2.3.4.4 viruses (H5N6 and H5N8) in Korea between 2016 and 2017. Wild birds spread clade 2.3.4.4.C H5N6 HPAI viruses from China to Japan and South Korea in late 2016, inflicting substantial economic losses on the poultry business,

particularly in Korea. In eight gene areas, emerging H5N6 viruses from 2017 shared 95 percent sequence similarity with previously reported Korean H5N6 viruses during 2016. Furthermore, in the 2016/2017 winter season, new H5N6 viruses shared 97 percent nucleotide similarities with three genes (PB2, PA and NP) with earlier Korean H5N8 viruses of clade 2.3.4.4.B. These findings suggest that novel H5N6 viruses are unlikely to be already circulating unnoticed in Korea, and hence they are more likely to be the consequence of a recent invasion facilitated by migrating birds [25].

In South Korea, there have been major HPAI outbreaks since 2003. The first outbreak in poultry farms resulted in a high-mortality epidemic [26,27]. This was followed by another one in 2006 and then in 2008. In 2014, a major HPAI outbreak belonging to clade 2.3.4.4 was first discovered in South Korean poultry farms, and it significantly progressed throughout several countries in Europe and Asia, resulting in the emergence of new H5NX subtypes [23]. Furthermore, between 2014 and 2018, major HPAI outbreaks in domestic poultry were also recorded. In South Korea, outbreaks caused by new reassortant clade 2.3.4.4 H5N6 viruses during the winter seasons of 2016–2017 and 2017–2018 resulted in the deaths of nearly one billion birds in 440 poultry farms [28].

It is highly critical to assess high-risk areas at various levels in order to properly implement preventive measures against H5NX clade 2.3.4.4 viruses.

Previous studies identified the transmission dynamics of HPAI infection spread, including the effects of changing the radius of culling birds around the infected zone on the local transmission of HPAI, although there was a gap in identifying the areas where the contributing factors in the disease transmission may cause severe HPAI outbreaks in terms of high bird infections and mortality in the future [21,29].

In this study, we focused on the avian influenza outbreaks that occurred in 2016–2017 in South Korean poultry farms. Hierarchical clustering on principal components analysis was performed on the data to identify clusters of regions as hotspots for the avian influenza epidemic in terms of a high number of susceptible birds, high infection, high point prevalence rate and high mortality in birds. The principal component analysis (PCA) approach, which is usually applied to reduce the dimensionality of the input, also delivers valuable insights into high dimensionality by deriving its fundamental properties. In our study, PCA served to extract useful features from the data, thereby making it convenient for interpretation [30].

Hierarchical clustering has been a powerful method for developing inbuilt classification systems [31]. One of its highly desirable characteristics is that it does not require the optimum number of clusters “K” to be specified in advance, as with other non-nested clustering techniques such as K-means, etc. Hierarchical methodologies, on the other hand, produce a unique nested hierarchy of clusters from which a partition can be obtained for any reasonable “number of clusters”. Consequently, hierarchical clustering offers an excellent framework for exploring relationships between clusters. The success of hierarchical clustering in a procedure can also be credited to dendrograms, an extremely insightful binary tree visualization of the clustering [32]. The analysis is performed in the R statistical computing and visualization environment, which is an accessible dialect of the S statistical computing language. It is free to use, runs on most operating systems and includes valuable contributions from leading computational experts in the field [33].

The results of the study highlight potential risk areas classified into significant clusters where HPAIV outbreaks may expand in the future due to a higher number of susceptible birds in the poultry farms at the herd level. The poultry farms with high infections and mortality in birds should implement proper disinfections inside and outside their surroundings. This study sheds light on the importance of implementing strict biosecurity measures in poultry farms with high flock density in order to avoid losses due to future epidemic threats.

## 2. Materials and Methods

### 2.1. Data Collection

The South Korean Animal and Plant Quarantine Agency (APQA) published the “Epidemiology report for outbreak of highly pathogenic avian influenza (HPAI) in Republic of Korea”, which included detailed information on poultry bird infections and death rates among susceptible birds in South Korean poultry farms. The dataset used in this study contained records from 2014 to 2018, covering variables such as the number of susceptible and infected birds, number of deaths and the number of culled birds across various farms. A total of 276 records were present in the dataset for analysis. Further detailed information regarding the dataset is available in the Supplementary Materials (Supplementary file.zip). The HPAI H5N6 epidemic began in November 2016 and lasted through March 2017, with a total of 343 cases. The epidemic occurred in a single phase for 107 days in total.

### 2.2. Hierarchical Clustering on Principal Components

The HCPC method combines 3 main multivariate data analysis methodologies, namely PCA, hierarchical clustering, and K-means. PCA is used to reduce the dimensionality of data by extracting continuous variables encompassing the most significant information. The findings of the PCA are then subjected to cluster analysis as a noise removal phase, which ensures robust clustering [34,35]. Finally, the K-means algorithm groups them into clusters. The poultry farms affected during the HPAI H5N6 epidemic were subjected to clustering analysis based on 5 variables, i.e., the number of infected birds, number of deaths in the infected birds, number of susceptible birds and number of culled birds.

### 2.3. Study Design and Setup

The R language for statistical computing was used for all of the analyses [36]. The R code for all the steps of our analysis is available in the Supplementary Materials (Supplementary file.zip). The avian influenza epidemic dataset “H5N6\_Data” was imported into R before starting the analysis. The packages required for effective hierarchical clustering analysis were installed and loaded into R in the first step. These packages included “FactoMineR” [37] and “factoextra” [38].

The second step of analysis involved feature extraction using PCA, a commonly used method to extract the hidden patterns in data. PCA is an unsupervised learning method that reduces ambiguity in large datasets while preserving trends over time. This was critical when working with our dataset, which contained variables that were correlated with one another. PCA divides the dataset into principal components (PCs), which are abstract representations of the data corresponding to new uncorrelated variables that successively maximize variance [39]. For our analysis, only the first three components were used by setting the option  $ncp = 3$ . As these three components were the most significant, the less significant components were ignored.

In the third step, hierarchical clustering on selected principal components was then applied. Using the “FactoMineR” package, the algorithm of the hierarchical clustering on principal components method was implemented. On the selected principal components, hierarchical clustering was computed by using the Ward criterion. The Ward criterion, which is based on multidimensional variance and is similar to principal component analysis, was implemented in the hierarchical clustering algorithm [34]. An initial partitioning was carried out through splitting the hierarchical tree. The partitioning was robustified with K-means clustering, and finally, the clusters were characterized with individuals and variables. Hierarchical clustering performs partitioning on two principles: (i) if individuals placed in the same class are close to each other (small within-class variability) and (ii) if individuals in different classes are far from each other (large between-class variability).

Mathematically, it can be expressed as Equation (1):

$$\underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (X_{iqk} - \bar{X}_k)^2}_{\text{total inertia}} = \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (X_{iqk} - \bar{X}_{qk})^2}_{\text{within-class inertia}} + \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (\bar{X}_{qk} - \bar{X}_k)^2}_{\text{between-class inertia}} \quad (1)$$

The above Equation (1) illustrates that the total inertia and the total variability can be broken down into within-class variability and between-class variability. Therefore,  $X_{iqk}$  is the value taken by  $i$ -th individual of the class  $q$  for the variable  $K$ . Moreover,  $\bar{X}_k$  is the mean of the  $k$ -th variable.  $\bar{X}_{qk}$  is the mean of  $k$  in class  $q$ , for the individuals in class  $q$ . The “within” inertia is the variability inside the class, which corresponds to the sum of the squared differences between the  $X_{iqk}$  and the  $\bar{X}_{qk}$ . The “between” inertia is the sum of the squared differences between the means of each class,  $\bar{X}_{qk}$ , and the means of each variable,  $\bar{X}_k$ . According to Huygens’ theorem, it is already known that total inertia is equal to the “between” inertia, plus the “within” inertia [40]. As a consequence, minimizing the “within” inertia is equivalent to maximizing the “between” inertia, because the total inertia is constant. Thus, we can either concentrate on minimizing the “within” inertia or maximizing the “between” inertia. These phenomena led us to put forward an index to measure the quality of a partition. Mathematically, partition quality is measured by:

$$0 \leq \frac{\text{between-class inertia}}{\text{total inertia}} \leq 1 \quad (2)$$

The above Equation (2) illustrates that the quality of the partition is the ratio of the “between” inertia over the total inertia. This ratio gives a value between 0 and 1, and the closer it is to 1, the better the partition is. The “between” inertia over the total inertia equals 0 when, for all variables  $k$ , the  $\bar{X}_{qk}$  are equal to  $\bar{X}_k$ , as shown in Equation (3):

$$\underbrace{\frac{\text{inertia}_{\text{between}}}{\text{inertia}_{\text{total}}}}_{=0} = 0 \Rightarrow \forall k, \forall q, \bar{X}_{qk} = \bar{X}_k \quad (3)$$

From Equation (3), it can be illustrated that all classes have the same means, for each variable. In this case, it is a partition that does not separate the classes and therefore does not help with clustering.

On the other hand, if the “between” inertia over the total inertia equals 1, this means that the “within” inertia is 0. This means that, inside each class, the individuals are identical. Thus, we have extremely homogenous classes, which is ideal for clustering. However, the result depends on the number of individuals and the number of classes. By increasing the number of classes, it is easier to have homogeneous ones. In the other extreme, if we have a small number of classes, the variability inside each class will be larger. Therefore, we consider the number of individuals and number of classes before making strong conclusions based on this criterion. This criterion for the quality of a partition suggests a new way to perform hierarchical clustering, known as Ward’s method [40]. The principles of Ward’s method are as follows:

1. Initialize : 1 class = 1 individual  $\Rightarrow$  Between – class inertia = total inertia
2. At each step : combine classes  $a$  and  $b$  that minimize the decrease in between – class inertia

According to Ward’s method, we start from the clustering where each individual is its own class, which means that, inside each class, there is no within-class variability, so the between-class variability must be equal to the total inertia, which leads to a perfect partition.

The aim of the Ward’s method is to choose two classes,  $a$  and  $b$ , so that their aggregation minimizes the decrease in the between-class inertia [40]. Essentially, the between-class

inertia can only decrease when we aggregate two classes, and we want to minimize this decrease.

Let us examine how we write the sum of the inertias of class “a” and class “b” as a function of the inertia of their aggregation.

$$Inertia(a) + Inertia(b) = Inertia(a \cup b) - \underbrace{\frac{m_a m_b}{m_a + m_b} d^2(a, b)}_{\text{to minimize}} \quad (4)$$

From Equation (4), it can be observed that the inertia of “a” plus “b” is equal to the inertia of the aggregation of these two classes, minus a certain quantity:  $(m_a m_b)$ , divided by  $(m_a + m_b)$ , multiplied by  $d^2(a, b)$ . Here,  $m_a$  is the number of individuals in the class “a”,  $m_b$  is the number of individuals in the class “b”, and  $d^2(a, b)$  is the squared distance between the centers of gravity of classes “a” and “b”. As we want the inertia of their union to be as close as possible to the inertia of “a”, plus the inertia of “b”, we need to minimize the second part of Equation (4), which contains two aspects: weights and the squared distance.

Looking at  $\frac{m_a m_b}{m_a + m_b}$ , this term intends to group together objects with small weights and avoid what are known as chain effects.

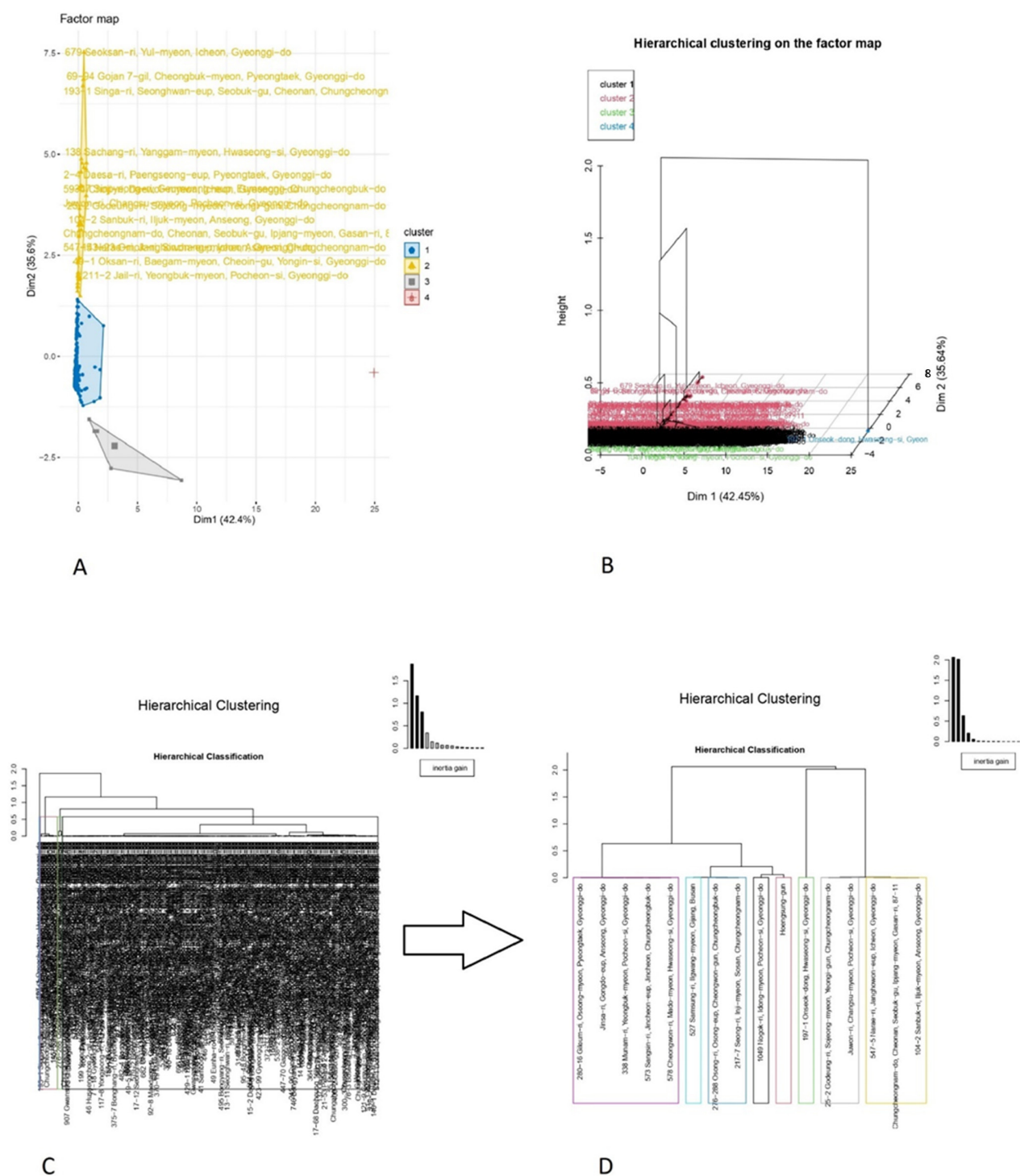
In the 4th step, the dendrogram created by hierarchical clustering was shown, using the “fviz\_dend()” function available in the “factoextra” package. Individuals on the principal component map were highlighted in the fifth step using the “factoextra” package’s function `fviz_cluster()` based on which cluster they were found to be associated with.

In the sixth step, using the `plot()` function, a three-dimensional plot combining hierarchical clustering with a factor map was created. The original data were generated, along with the cluster designations and quantitative characteristics that best defined each cluster. Important information regarding all clusters, including the “Mean in Category”, “Overall Mean”, “p value” and principal dimensions, significantly associated with these clusters were calculated. Finally, a list of the most key individuals (poultry farms) in each cluster was created.

### 3. Results and Discussion

This section provides details regarding all the analyses performed under this study. The analysis results have mostly been displayed as visualizations for easy understanding. The main purpose of this analysis is to extract the hidden patterns in the data and conveniently display them as appropriate visualizations.

The output of the hierarchical clustering algorithm on principal components resulted in a “factor map”, “hierarchical clustering on the factor map” and “dendrogram”. These visualizations can be observed in Figure 1. The hierarchical cluster tree showing four clusters in black, orange, green and blue color triangles is presented in Figure 1.



**Figure 1.** Panel (A) depicts the PCA plot showing poultry farms assigned to each of the four clusters. Cluster 1 and 2 have coordinates on dimension 2, which explain 35.6% of the variance in the data. Cluster 3 has coordinates on dimension 3, which explain 17.33% of the variance. Cluster 4 has coordinates on dimension 1, which explain 42.4% of the variance in the data. Panel (B) shows a combined plot of hierarchical clustering and PCA in a three-dimensional view. Panel (C) shows the hierarchical cluster tree, which has been cut to a height of  $<0.5$ , showing the cluster dendrogram dividing the poultry farms into 4 clusters along with a graphical output showing the loss of inertia. Cluster 1 is enclosed in the black rectangle, whereas the orange rectangle encloses the areas coming under Cluster 2. The green and blue triangles enclose Clusters 3 and 4, respectively. Panel (D) shows the most important elements of the clusters that are closest to the cluster centers.

Figure 1 shows maps with identified clusters.

The individuals belonging to each cluster separated in Figure 1C can be observed in the table available in the Supplementary Materials (supplementary.zip). For clear visualization, the most important individuals belonging to each cluster, and which are nearest to the centers of their respective cluster, are shown in Figure 1D.

In Figure 1C, we present the cluster dendrogram obtained with a graphical output showing the loss of inertia. At the first bar in the graph, we see that there is a large loss in inertia when passing from two classes to one, which means that it is not a good option to group them together. In the same way, we see that there is a loss of inertia when passing from three classes to two, and from four classes to three. In contrast, there is very little loss of inertia when passing from five classes to four.

Figure 1D, included alongside Figure 1C, shows the most important variables belonging to different clusters. Figure 1D shows eight clusters due to the lower number of individuals, which leads to more partitions. It is simpler to have homogenous classes as the number of classes is increased. At the opposite extreme, if we have a small number of classes, the diversity within each class will be greater [41]. However, if we cut the tree at a height of 0.5, we will observe the same four groups (see Figure 1D).

Eigenvalues obtained through PCA are used to determine the number of PCs to keep, such that the most significant features are retained, and the trivial ones are ignored. The eigenvalues presented in Table 1 show that a total of five PCs were obtained. The eigenvalue of the first principal component explained 42.45% of the variation, while 35.5% of the variation is explained by the second eigenvalue. The third eigenvalue explained 17.33% of the variance. The fourth eigenvalue explained 4.58% of the variation. The fifth eigenvalue explained almost no variance. The cumulative percentage explained is calculated by adding the running total of the successive proportions of variation explained by each principal component; results can be observed in Table 1.

**Table 1.** Five principal components with their corresponding eigenvalues, percentage of variance and cumulative percentage of variance for each principal component.

| Principal Component | Eigenvalue                | Percentage of Variance | Cumulative Percentage of Variance |
|---------------------|---------------------------|------------------------|-----------------------------------|
| Component 1         | $2.122466 \times 10^0$    | 42.44932               | 42.44932                          |
| Component 2         | $1.781823 \times 10^0$    | 35.63646               | 78.08578                          |
| Component 3         | $8.667983 \times 10^{-1}$ | 17.33597               | 95.42175                          |
| Component 4         | $2.289126 \times 10^{-1}$ | 4.578253               | 100.00000                         |
| Component 5         | $2.225099 \times 10^{-8}$ | 0.0000004              | 100.00000                         |

The covariance matrix is symmetric, and symmetric matrices include orthogonal eigenvectors; the PCA produced orthogonal components [35,42–44].

The first principal component was the eigenvector with the highest eigenvalue  $2.122466 \times 10^0$ ; this component explained the most variation in the data, making it the most significant feature of the data. Similarly, the eigenvector with the second largest eigenvalue,  $1.781823 \times 10^0$ , was the second principal component, and the third and fourth components were also found to be significant. However, the fifth principal component explained almost no variance. Hence, this feature was not used in further analyses.

After applying hierarchical clustering on PCs, the combined plot of hierarchical clustering and PCA was created, which can be seen in Figure 1.

The analysis of the clusters revealed that 308 poultry farms were assigned to Cluster 1, 36 poultry farms were assigned to Cluster 2, five poultry farms were assigned to Cluster 3, and one poultry farm was assigned to Cluster 4. The data with cluster assignments are provided in the Supplementary Materials for further analysis and observation (Supplementary file.zip).

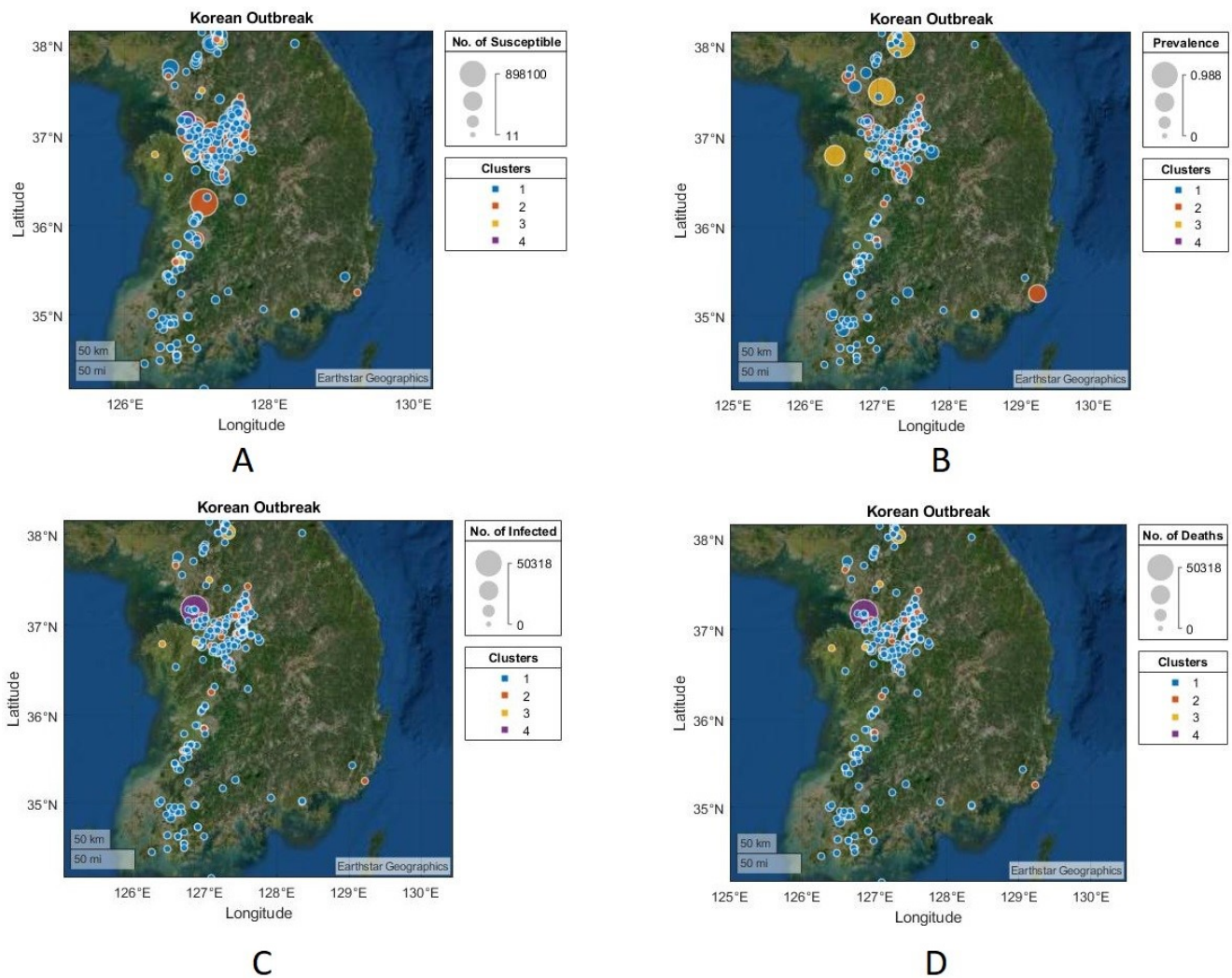
Each cluster was found characterized by the variables (numerical characteristics) in our dataset. The numerical characteristics that best define each cluster are displayed in Table 2.

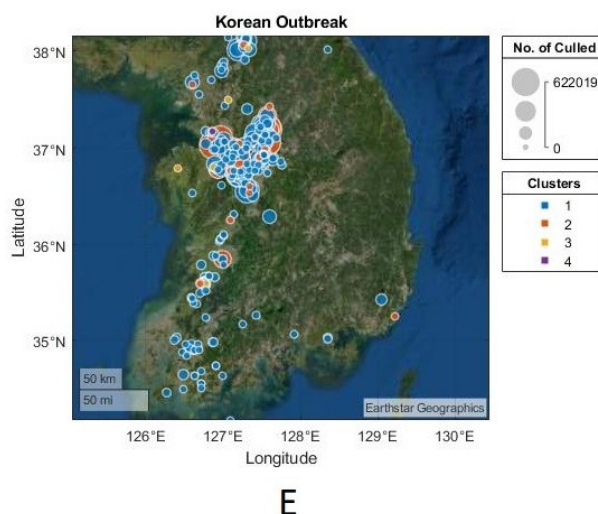
**Table 2.** The clusters characterized by the variables.

| Clusters  | Variables                      | V.Test     | Mean in Category | Overall Mean | Sd in Category            | Overall Sd                | p Value                    |
|-----------|--------------------------------|------------|------------------|--------------|---------------------------|---------------------------|----------------------------|
| Cluster 1 | No_of_infected_birds           | −3.241110  | 137.3366         | 314.8319     | $4.587386 \times 10^2$    | $2.778956 \times 10^3$    | $1.190651 \times 10^{-3}$  |
|           | No_of_deaths_in_infected_birds | −3.241184  | 137.3010         | 314.8006     | $4.587482 \times 10^2$    | $2.778960 \times 10^3$    | $1.190342 \times 10^{-3}$  |
|           | Point_prevalence               | −5.200421  | 0.0084           | 0.01728519   | $2.686831 \times 10^{-2}$ | $8.657319 \times 10^{-2}$ | $1.988379 \times 10^{-7}$  |
|           | No_of_culled_birds             | −13.262405 | 29,457.82        | 54,353.69    | $3.824085 \times 10^4$    | $9.525626 \times 10^4$    | $3.824311 \times 10^{-40}$ |
| Cluster 2 | No_of_susceptible_birds        | −13.526878 | 38,400.67        | 68,965.80    | $4.663563 \times 10^4$    | $1.146614 \times 10^5$    | $1.085365 \times 10^{-41}$ |
|           | No_of_culled_birds             | 14.78849   | 277,088.1        | 54,353.69    | 143,163.3                 | 95,256.26                 | $1.738220 \times 10^{-49}$ |
| Cluster 3 | No_of_susceptible_birds        | 14.64812   | 334,529.9        | 68,965.80    | 170,478.1                 | 114,661.44                | $1.384946 \times 10^{-48}$ |
| Cluster 4 | Point_prevalence               | 16.56209   | 0.65484          | 0.01728519   | 0.2511564                 | 0.08657319                | $1.309701 \times 10^{-61}$ |
|           | No_of_infected_birds           | 17.99351   | 50,318           | 314.8319     | 0                         | 2778.956                  | $2.190528 \times 10^{-72}$ |
|           | No_of_deaths_in_infected_birds | 17.99349   | 50,318           | 314.8006     | 0                         | 2778.960                  | $2.190949 \times 10^{-72}$ |

Table 2 shows quantitative variables significantly associated with clusters. When compared to other clusters, poultry farms in Cluster 4 had significantly higher average bird infections with high mortality (mean in category). Similarly, poultry farms in Cluster 2 had a higher average number of susceptible birds and number of culled birds (mean in category) as compared to other clusters. When compared to other clusters, poultry farms in Cluster 3 had a significantly higher average point prevalence rate of HPAI H5N6 cases (mean in category).

Each cluster with numerical characteristics was also displayed on maps, as shown in Figure 2, for convenient interpretation of the analysis results.

**Figure 2.** Cont.



**Figure 2.** Panel (A) presents the number of susceptible birds in each of the 4 clusters. From the figure, it can be observed that Cluster 2 had highest number of susceptible birds, followed by Cluster 1. Panel (B) shows the prevalence rate of HPAI infections. From the figure, it can be observed that Cluster 2 had the highest prevalence rate of infections, followed by Cluster 2 and then Cluster 1. Panel (C) indicates the number of infected birds in each of the 4 clusters. The purple color indicates that Cluster 4 had the highest number of infected birds, followed by the yellow-colored Cluster 3. Panel (D) shows the number of deaths in infected birds in each of the 4 clusters. Cluster 4 shows a severe mortality rate, followed by Cluster 3. Panel (E) shows the number of culled birds in each of the 4 clusters. Cluster 2, followed by Cluster 1, can be seen with the highest numbers of culled birds.

Figure 2 shows maps of South Korea, presenting the intensity of the variables.

These four significant clusters comprising a number of poultry farms had coordinates on axes 1, 2 and 3 on the factor map obtained as a result of PCA. Table 3 shows the principal dimensions belonging to each cluster.

**Table 3.** Clusters with their respective coordinates.

| Clusters  | Variables   | V.Test     | Mean in Category | Overall Mean                | Sd in Category | Overall Sd | p Value                    |
|-----------|-------------|------------|------------------|-----------------------------|----------------|------------|----------------------------|
| Cluster 1 | Dimension 1 | −5.205785  | −0.1494576       | $-1.446591 \times 10^{-16}$ | 0.2813545      | 1.4568686  | $1.931785 \times 10^{-7}$  |
|           | Dimension 3 | −7.215580  | −0.1323859       | $2.246838 \times 10^{-16}$  | 0.2689891      | 0.9310201  | $5.370475 \times 10^{-13}$ |
|           | Dimension 2 | −13.067666 | −0.3437492       | $2.938256 \times 10^{-16}$  | 0.5255536      | 1.3348494  | $5.038794 \times 10^{-39}$ |
| Cluster 2 | Dimension 2 | 15.493287  | 3.2699856        | $2.938256 \times 10^{-16}$  | 1.5995139      | 1.3348494  | $3.850870 \times 10^{-54}$ |
|           | Dimension 3 | 2.688344   | 0.3957436        | $2.246838 \times 10^{-16}$  | 0.2777148      | 0.9310201  | $7.180736 \times 10^{-3}$  |
| Cluster 3 | Dimension 3 | 15.613585  | 6.463693         | $2.246838 \times 10^{-16}$  | 2.3301834      | 0.9310201  | $5.883553 \times 10^{-55}$ |
|           | Dimension 1 | 4.729651   | 3.063857         | $-1.446591 \times 10^{-16}$ | 2.8833684      | 1.4568686  | $2.249064 \times 10^{-6}$  |
| Cluster 4 | Dimension 2 | −3.739145  | −2.219339        | $2.938256 \times 10^{-16}$  | 0.5871431      | 1.3348494  | $1.846473 \times 10^{-4}$  |
|           | Dimension 1 | 17.124859  | 24.948669        | $-1.446591 \times 10^{-16}$ | 0              | 1.4568686  | $9.684419 \times 10^{-66}$ |
|           | Dimension 3 | −6.077204  | −5.657999        | $2.246838 \times 10^{-16}$  | 0              | 0.9310201  | $1.222959 \times 10^{-9}$  |

Table 3 output indicates that Cluster 1 has lower coordinates on dimension 2. Cluster 2 has high coordinates on dimension 2. Cluster 3 has high coordinates on dimension 3. Similarly, Cluster 4 has high coordinates on dimension 1.

Hierarchical clustering is based on an appropriate metric as a measure of distance between pairs of observations [45]. Cursory scrutiny of the hierarchical clustering dendrogram, which serves as a summary of the distance matrix, reveals the similarities and differences between individuals in each cluster. From Figure 1D, it can be observed that poultry farm “25-2 Godeung-ri, Sojeong-myeon, Yeongi-gun, Chungcheongnam-do” is closer to “104-2 Sanbuk-ri, Iljuk-myeon, Anseong, Gyeonggi-do” as compared to “197-1 Onseok-dong, Hwaseong-si, Gyeonggi-do”. Similarly, poultry farm “527 Samsung-ri, Ilgwang-myeon, Gijang, Busan” is closer to “Hoengsung-gun” rather than “578 Cheongwon-ri, Mado-myeon, Hwaseong-si, Gyeonggi-do”.

The individuals lying on the same branch of the hierarchical dendrogram tree are closer to one another and closest to their respective clusters' centers. The same phenomena are displayed in Table 4, denoted as paragons. The individuals in each cluster obtained in Figure 1D are similar to the individuals shown in Table 4 as paragons of each cluster. Poultry farm "578 Cheongwon-ri, Mado-myeon, Hwaseong-si, Gyeonggi-do" belongs to Cluster 1 and is the closest to the center of the first cluster. Poultry farm "104-2 Sanbuk-ri, Iljuk-myeon, Anseong, Gyeonggi-do" belongs to Cluster 2 and is the closest to the center of the second cluster. Poultry farm "276-288 Osong-ri, Osong-eup, Cheongwon-gun, Chungcheongbuk-do" belongs to Cluster 3 and is closest to the center of the third cluster. Poultry farm "197-1 Onseok-dong, Hwaseong-si, Gyeonggi-do" belongs to Cluster 4 and is closest to the center of the fourth cluster (see Table 4). The individuals, based on the dissimilarity matrix, have been separated into different clusters, and the individuals belonging to one cluster are at a farther distance from the individuals belonging to another cluster. The phenomenon is expressed in Table 5. Thus, these poultry farms were the most important ones characterized by high infections, birds' mortality as well as a high point prevalence rate of the HPAI H5N6 epidemic. These poultry farms should be given greater priority while implementing adequate biosecurity measures to limit the occurrence and spread of epidemics in the future. The locations of the poultry farms belonging to these clusters were compared with the maps downloaded from online sources, showing the domestic duck density and domestic chicken density in the provinces of South Korea, which can be found in Figures 3 and 4 [46,47].

**Table 4.** The most important variables of each cluster, closest to their respective cluster centers.

| Clusters  | Poultry Farms                                                        | Paragons   |
|-----------|----------------------------------------------------------------------|------------|
| Cluster 1 | 578 Cheongwon-ri, Mado-myeon, Hwaseong-si, Gyeonggi-do               | 0.05777456 |
|           | 573 Sangsin-ri, Jincheon-eup, Jincheon, Chungcheongbuk-do            | 0.05810394 |
|           | Jinsa-ri, Gongdo-eup, Anseong, Gyeonggi-do                           | 0.06201014 |
|           | 338 Munam-ri, Yeongbuk-myeon, Pocheon-si, Gyeonggi-do                | 0.06314534 |
|           | 280-16 Gileum-ri, Oseong-myeon, Pyeongtaek, Gyeonggi-do              | 0.06798525 |
| Cluster 2 | 104-2 Sanbuk-ri, Iljuk-myeon, Anseong, Gyeonggi-do                   | 0.1651389  |
|           | Chungcheongnam-do, Cheonan, Seobuk-gu, Ipjang-myeon, Gasan-ri, 87-11 | 0.1969258  |
|           | 25-2 Godeung-ri, Sojeong-myeon, Yeongi-gun, Chungcheongnam-do        | 0.2072500  |
|           | 547-5 Narae-ri, Janghowon-eup, Icheon, Gyeonggi-do                   | 0.3301516  |
|           | Juwon-ri, Changsu-myeon, Pocheon-si, Gyeonggi-do                     | 0.4044791  |
| Cluster 3 | 276-288 Osong-ri, Osong-eup, Cheongwon-gun, Chungcheongbuk-do        | 2.051024   |
|           | 217-7 Seong-ri, Inji-myeon, Sosan, Chungcheongnam-do                 | 2.183701   |
|           | Hoengsung-gun                                                        | 3.337730   |
| Cluster 4 | 527 Samsung-ri, Ilgwang-myeon, Gijang, Busan                         | 3.591180   |
|           | 1049 Nogok-ri, Idong-myeon, Pocheon-si, Gyeonggi-do                  | 6.117946   |
|           | 197-1 Onseok-dong, Hwaseong-si, Gyeonggi-do                          | 0          |

Table 4 shows paragons. The top five individuals as the regions closest to the cluster center are presented for each cluster, with the exception of Cluster 4, which contains a single poultry farm.

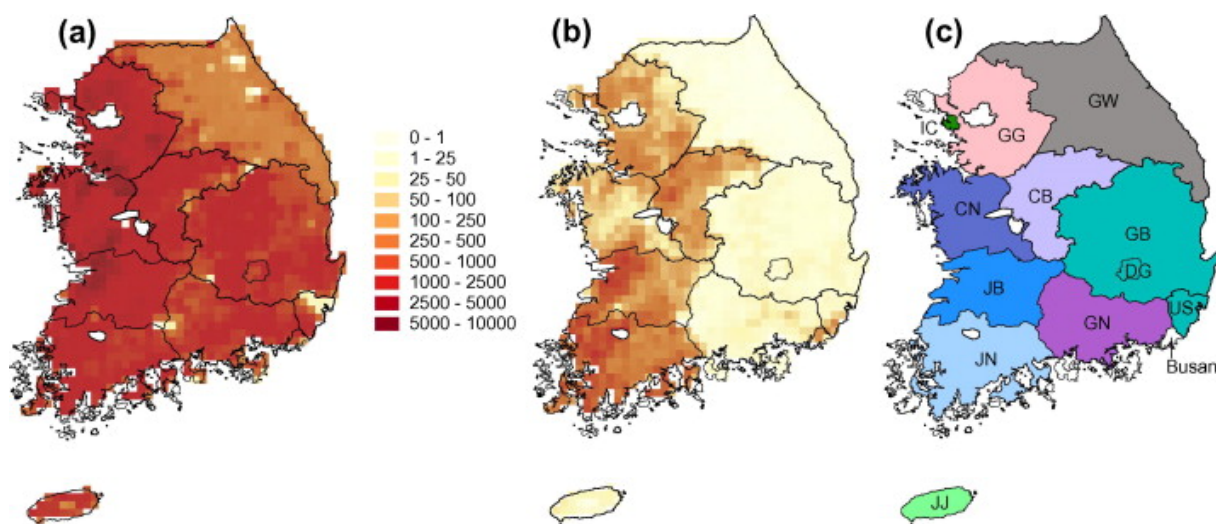
**Table 5.** The individuals separated based on dissimilarity matrix.

| Clusters  | Poultry Farms                                          | Distance |
|-----------|--------------------------------------------------------|----------|
| Cluster 1 | 911 Bangcho-ri, Iljuk-myeon, Anseong, Gyeonggi-do      | 4.767073 |
|           | 280-3 Aseong-ri, Sosu-myeon, Goesan, Chungcheongbuk-do | 4.697875 |
|           | RGPH+F52 Muan-gun, Jeollanam-do                        | 4.630825 |
|           | 75 Chowonji-ri, Daegot-myeon, Gimpo-si, Gyeonggi-do    | 4.568716 |
|           | 75-1 Gongchon-dong, Seo-gu, Incheon                    | 4.545588 |

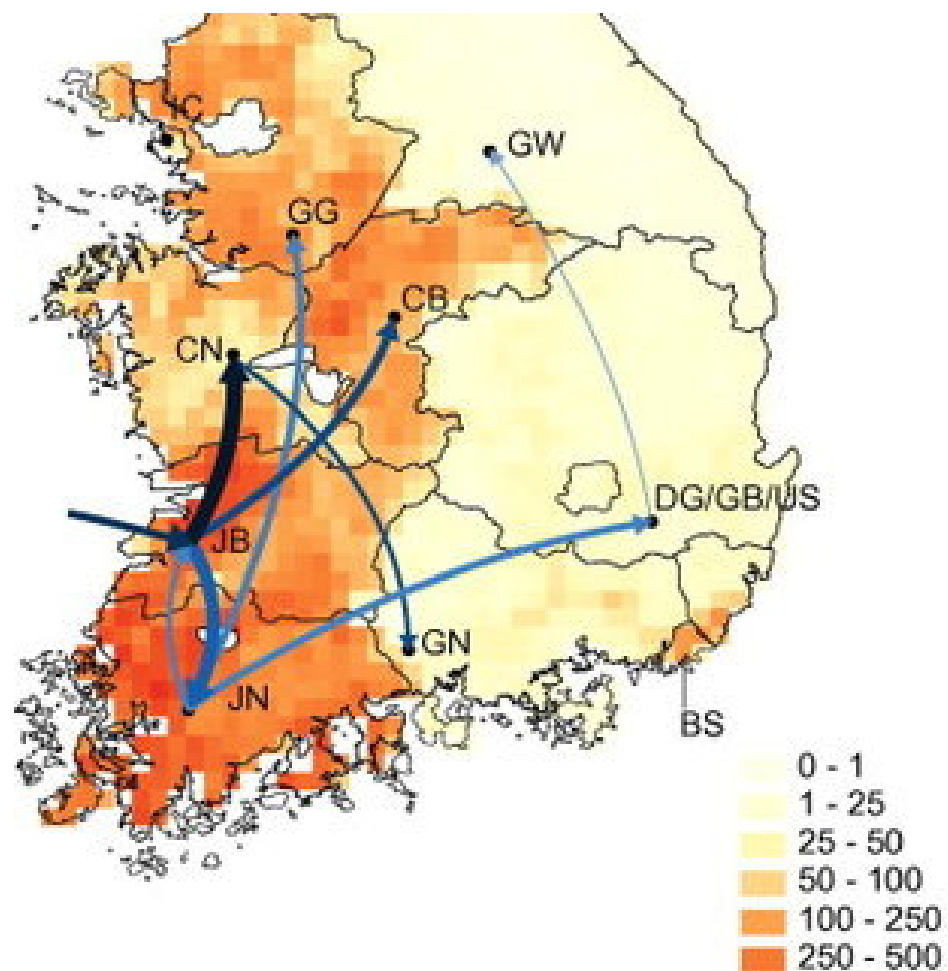
Table 5. Cont.

| Clusters  | Poultry Farms                                                        | Distance  |
|-----------|----------------------------------------------------------------------|-----------|
| Cluster 2 | 679 Seoksan-ri, Yul-myeon, Icheon, Gyeonggi-do                       | 8.004391  |
|           | 69-94 Gojan 7-gil, Cheongbuk-myeon, Pyeongtaek, Gyeonggi-do          | 7.313310  |
|           | 193-1 Singa-ri, Seonghwan-eup, Seobuk-gu, Cheonan, Chungcheongnam-do | 7.152351  |
|           | 138 Sachang-ri, Yanggam-myeon, Hwaseong-si, Gyeonggi-do              | 5.309652  |
|           | Gimdae 1-gil, Ganam-myeon, Yeosu-gun, Gyeonggi-do                    | 5.240544  |
| Cluster 3 | 1049 Nogok-ri, Idong-myeon, Pocheon-si, Gyeonggi-do                  | 12.786330 |
|           | Hoengsung-gun                                                        | 10.572264 |
|           | 276-288 Osong-ri, Osong-eup, Cheongwon-gun, Chungcheongbuk-do        | 5.697607  |
|           | 217-7 Seong-ri, Inji-myeon, Sosan, Chungcheongnam-do                 | 5.695346  |
| Cluster 4 | 527 Samsung-ri, Ilgwang-myeon, Gijang, Busan                         | 4.127495  |
|           | 197-1 Onseok-dong, Hwaseong-si, Gyeonggi-do                          | 25.08336  |

Table 5 displays the distances between poultry farms in one cluster and poultry farms in other clusters. Poultry farm “911 Bangcho-ri, Iljuk-myeon, Anseong, Gyeonggi-do” belongs to Cluster 1 and is far away from the centers of Clusters 2, 3 and 4. Poultry farm “679 Seoksan-ri, Yul-myeon, Icheon, Gyeonggi-do” belongs to Cluster 2 and is far away from the centers of Clusters 1, 3 and 4. Poultry farm “1049 Nogok-ri, Idong-myeon, Pocheon-si, Gyeonggi-do” belongs to Cluster 3 and is far away from the centers of Clusters 1, 2 and 4. Poultry farm “197-1 Onseok-dong, Hwaseong-si, Gyeonggi-do” belongs to Cluster 4 and is far away from the centers of Clusters 1, 2 and 3.



**Figure 3.** Maps showing domestic poultry density (number per kilometer, colors in key) in ROK according to the Gridded Livestock of the World 2.0 (Robinson et al., 2014). (a) Domestic chicken density. (b) Domestic duck density. (c) Map of provinces. Colors correspond to the branch color scheme used in Figure 3. Province abbreviations are as follows: CB: Chungbuk, CN: Chungnam, DG: Daegu, GB: Gyeongbuk, GG: Gyeonggi, GN: Gyeongnam, GW: Gangwon, IC: Incheon, JB: Jeonbuk, JN: Jeonnam, JJ: Jeju, JN: Jeonnam, US: Ulsan. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article (accessed on 24 February 2022) [46,47]).



**Figure 4.** Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea, according to the Gridded Livestock of the World 2.0 (Robinson et al., 2014). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article (accessed on 24 February 2022) [46,47]).

Figure 3 showing domestic chicken density and domestic duck density in the provinces of South Korea.

Figure 4 displays the migratory trajectories of viral transmission via migratory waterfowls.

In this study, multivariate statistical approaches and hierarchical cluster analysis were used to assess epidemic risks in South Korean regions having HPAI epidemic cases. HCPC allowed us to construct hierarchies of similarly HPAI-affected regions.

The HCPC methodology offers acceptable clustering based on a set of epidemic characteristics. Clustering involves calculating the distance between each pair of objects (for example, poultry farms' locations) in terms of the study's specified variables (e.g., number of infected, number of culled, number of susceptible, number of deaths, prevalence), and then grouping objects that are closest enough. Cluster analysis resulted in a comparison and evaluation of HPAI epidemic cases across regions, resulting in the identification of clusters of regions with comparable data. The first benefit of our method is that it uses an objective clustering approach to investigate PCA results, resulting in a better cluster arrangement [35]. At the end of the study, according to the HPAI indicators, regions were divided into four subgroups as high-risk areas in terms of the HPAI epidemic.

From the results of Table 2 and Figure 2, it is demonstrated that the average number of infected birds, as well as the mortality in affected birds, in poultry farms belonging to Cluster 4 was higher (50,318) than the overall mean (314.8) across all clusters. The 197-1 Onseok-dong,

Hwaseong-si, Gyeonggi-do poultry farm was found associated with Cluster 4. It can be argued that Cluster 4 is distinguished by a high number of infected birds as well as mortality.

Furthermore, the poultry farms in Cluster 3 had a higher average point prevalence rate (0.654) compared to other clusters. Consequently, it was discovered that poultry farms in Cluster 3 had a relatively higher point prevalence rate. The poultry farms 276-288 Osong-ri, Osong-eup, Cheongwon-gun, Chungcheongbuk-do, 217-7 Seong-ri, Inji-myeon, Sosan, Chungcheongnam-do, Hoengsung-gun, Gangwon-do, 527 Samsung-ri, Ilgwang-myeon, Gijang, Busan, 1049 Nogok-ri, Idong-myeon, Pocheon-si, Gyeonggi-do were included in Cluster 3. Poultry farms in Cluster 2 had the highest average number of susceptible birds (334,529.9) among all clusters. Similarly, the average number of culled birds (277,088.1) was higher in Cluster 2 across all clusters.

It is clear that poultry farms belonging to Cluster 2 are distinguished by a greater number of susceptible as well as culled birds. According to these findings, the poultry farms in Cluster 2 had a high rate of culling birds in order to prevent further spread and mortality in a large number of susceptible birds. Poultry farms in Cluster 2 that were nearest to the cluster center were: 104-2 Sanbuk-ri, Iljuk-myeon, Anseong, Gyeonggi-do, Chungcheongnam-do, Cheonan, Seobuk-gu, Ipjang-myeon, Gasan-ri, 87-11, 25-2 Godeung-ri, Sojeong-myeon, Yeongi-gun, Chungcheongnam-do, 547-5 Narae-ri, Janghowon-eup, Icheon, Gyeonggi-do, Juwon-ri, Changsu-myeon, Pocheon-si, Gyeonggi-do.

Research studies have shown that the proactive culling of poultry birds can control the spread of HPAI epidemics causing high infections as well as mortality [21]. Poultry farms in Cluster 1 were found characterized by a large susceptible bird population, infection rate, death rate and number of culled birds, but the variables were found with negative T statistics, and the mean values of the variables were found to be lower than the average mean across all clusters (see Table 2). Poultry farms included in Cluster 1 that were closest to the cluster center were: 578 Cheongwon-ri, Mado-myeon, Hwaseong-si, Gyeonggi-do, 573 Sangsin-ri, Jincheon-eup, Jincheon, Chungcheongbuk-do, Jinsa-ri, Gongdo-eup, Anseong, Gyeonggi-do, 338, Munam-ri, Yeongbuk-myeon, Pocheon-si, Gyeonggi-do, 280-16 Gileum-ri, Oseong-myeon, Pyeongtaek, Gyeonggi-do. Poultry farms in various provinces were identified to form clusters considered hotspots, e.g., Gyeonggi-do, Chungcheongnam-do, Chungcheongbuk-do, Gangwon-do and Busan.

Referring to Figures 3 and 4 from an online source article portraying the movements of migratory birds that might be hypothesized as plausible routes of conveyance for the HPAI H5N6 outbreak, comparable locations were identified in our analysis as a result of the HCPC approach [46,47].

According to Hill, S.C., et al. [46,47], Gyeonggi-do, Chungcheongnam-do, Chungcheongbuk-do, Busan, except Gangwon province, have a higher duck population density as well as chicken density. Duck poultry normally do not show any clinical indications of HPAI infection; rather, these birds facilitate the spread of HPAI diseases into the chicken population [21].

Therefore, it can be hypothesized that duck populations cause the occurrence and spread in the regions identified as hotspots, where these birds caused significant infections and mortality in chickens during the H5N6 epidemic. In the case of Gangwon province, evaluating online Figure 4, it is possible to see possible routes of migratory birds transporting the virus from Jeollanam-do (a province with high duck and chicken density) to Daegu, Gyeongbuk, Ulsan and then to Gangwon province, which was identified as a hotspot based on our findings. It is evident that our findings support the prior study's conclusions [46,47]. Understanding the risk factors for outbreaks, especially clustering, has important implications for disease control and prevention [48–50].

A high number of susceptible birds (high flock density) can be a risk factor for the severity of HPAI epidemics in the poultry farms belonging to Clusters 1 and 2. Tiensin et al. [51] demonstrated that flock density is the key risk factor for AIV propagation as well as prevalence. Direct contact with secretions, excrement and aerosol droplets can disseminate AIV from bird to bird. As a result, increased flock density can hasten the transmission of AIV from bird to bird [52].

As a consequence, there is a substantial relationship between flock density and AIV transmission in poultry. In addition, because of the higher point prevalence rate, poultry farms in Cluster 3 should be investigated further in terms of potential risk factors and biosecurity measures. To avoid AIV spread, strict farm biosecurity practices in the case of higher flock density in poultry farms are recommended to limit the risks of future HPAI epidemics [30]. The spread of AIV might be exacerbated by a mix of different types of flocks or inadequate hygienic conditions in chicken nesting areas [53].

Due to the unavailability of related data, there were some limitations to this study: the density of poultry farms, wild birds as well as land covered by water bodies in each region of the HPAI outbreak cases were not analyzed using multivariable clustering analysis; however, these characteristics contribute to a greater risk of an HPAI outbreak [54,55]. In-depth multivariable clustering analysis of HPAI epidemics with space–time interaction will be necessary in the future, taking into account wild bird density and poultry density at the district level.

#### 4. Conclusions

Finally, the results indicate the severity of HPAI epidemic cases in the poultry farms belonging to Cluster 3 with a high prevalence rate of HPAI H5N6, while Cluster 4 was linked to a substantial poultry infection and mortality rate despite the fact that there was not a higher number of susceptible birds in comparison. High-level biosecurity practices should be implemented in the poultry farms belonging to Clusters 3 and 4. Poultry farms in Cluster 2 were more vulnerable to HPAI infections due to a larger number of susceptible birds in the herd, including other potential risk factors, which should be investigated in the future.

Our analysis provides authorities with a regional view of the H5N6 emergence in South Korean poultry farms. From the overview of the results, it can be concluded that poultry farms with high numbers of susceptible birds at an individual poultry farm should implement strategies to avoid severe outbreaks leading to massive bird mortality as well as culling. Separating flocks into different chicken houses, rather than having a high density of flocks in a single poultry house, can be a helpful technique to reduce HPAI infections; however, appropriate precautionary measures should be applied at each poultry house's entrance.

It is crucial to distinguish between distinct clusters of epidemic characters in order to evaluate the effectiveness of the South Korean government's regulations. This demarcation may also aid in directing future countermeasures, which should be implemented in accordance with area conditions. In addition, making suitable modifications to our strategy might help us to manage emergencies in the future.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/sym14030598/s1>.

**Author Contributions:** Conceptualization: J.-i.L., G.-H.S., C.-M.L.; data curation: S.A., K.-Y.K.; formal analysis: S.A., K.-Y.K.; investigation: S.A., K.-Y.K., J.-i.L.; methodology: S.A., K.-Y.K.; resources: S.A., K.-Y.K., J.-i.L., G.-H.S., C.-M.L.; software: S.A., K.-Y.K.; supervision: J.-i.L., G.-H.S., C.-M.L.; validation: S.A., K.-Y.K., J.-i.L., G.-H.S., C.-M.L.; visualization: S.A., K.-Y.K.; writing—original draft: S.A., K.-Y.K.; writing—review and editing: S.A., K.-Y.K., J.-i.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Korea Smart Farm R & D Foundation (KosFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Ministry of Science and ICT (MSIT), Rural Development Administration (RDA) (No. 421015-04).

**Data Availability Statement:** The datasets that corroborate the conclusions of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** We want to thank all the members of the “Animal and Plant Quarantine Agency, Korea”, especially Daesung Yoo, for providing us with the data to perform the work successfully. We also appreciate the cooperation of the poultry farm experts to guide us in collecting the information.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Fidler, D.P. *Global Outbreak of Avian Influenza A (H5N1) and International Law*; ASIL Insights (January 2004); The American Society of International Law: Washington, DC, USA, 2004.
2. Webby, R.; Webster, R. Emergence of influenza A viruses. *Philos. Trans. R. Soc. Lond. Ser. B* **2001**, *356*, 1817–1828. [[CrossRef](#)] [[PubMed](#)]
3. Webster, R.G.; Bean, W.J.; Gorman, O.T.; Chambers, T.M.; Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **1992**, *56*, 152–179. [[CrossRef](#)] [[PubMed](#)]
4. Olsen, B.; Munster, V.J.; Wallensten, A.; Waldenström, J.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Global patterns of influenza A virus in wild birds. *Science* **2006**, *312*, 384–388. [[CrossRef](#)]
5. Krammer, F.; Palese, P. Influenza virus hemagglutinin stalk-based antibodies and vaccines. *Curr. Opin. Virol.* **2013**, *3*, 521–530. [[CrossRef](#)] [[PubMed](#)]
6. Chatziprodromidou, I.P.; Arvanitidou, M.; Guitian, J.; Apostolou, T.; Vantarakis, G.; Vantarakis, A. Global avian influenza outbreaks 2010–2016: A systematic review of their distribution, avian species and virus subtype. *Syst. Rev.* **2018**, *7*, 17. [[CrossRef](#)]
7. Webster, R.G. Antigenic variation in influenza viruses. In *Origin and Evolution of Viruses*; Academic Press: Cambridge, MA, USA, 1999; pp. 377–390.
8. Fiebig, L.; Soyka, J.; Buda, S.; Buchholz, U.; Dehnert, M.; Haas, W. Avian influenza A (H5N1) in humans: New insights from a line list of World Health Organization confirmed cases, September 2006 to August 2010. *Eurosurveillance* **2011**, *16*, 19941. [[CrossRef](#)] [[PubMed](#)]
9. Djunaidi, H.; Djunaidi, A.C. The economic impacts of avian influenza on world poultry trade and the US poultry industry: A spatial equilibrium analysis. *J. Agric. Appl. Econ.* **2007**, *39*, 313–323. [[CrossRef](#)]
10. Bui, C.; Bethmont, A.; Chughtai, A.; Gardner, L.; Sarkar, S.; Hassan, S.; Seale, H.; MacIntyre, C.R. A systematic review of the comparative epidemiology of avian and human influenza A H5N1 and H7N9—lessons and unanswered questions. *Transbound. Emerg. Dis.* **2016**, *63*, 602–620. [[CrossRef](#)]
11. Abolnik, C. A current review of avian influenza in pigeons and doves (Columbidae). *Veter. Microbiol.* **2014**, *170*, 181–196. [[CrossRef](#)] [[PubMed](#)]
12. Tong, S.; Li, Y.; Rivailler, P.; Conrardy, C.; Castillo, D.A.A.; Chen, L.-M.; Recuenco, S.; Ellison, J.A.; Davis, C.T.; York, I.A.; et al. A distinct lineage of influenza A virus from bats. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 4269–4274. [[CrossRef](#)]
13. Tong, S.; Zhu, X.; Li, Y.; Shi, M.; Zhang, J.; Bourgeois, M.; Yang, H.; Chen, X.; Recuenco, S.; Gomez, J.; et al. New world bats harbor diverse influenza A viruses. *PLoS Pathog.* **2013**, *9*, e1003657. [[CrossRef](#)] [[PubMed](#)]
14. Tumpey, M.T.; Kapczynski, D.R.; Swayne, D.E. Comparative susceptibility of chickens and turkeys to avian influenza A H7N2 virus infection and protective efficacy of a commercial avian influenza H7N2 virus vaccine. *Avian Dis.* **2004**, *48*, 167–176. [[CrossRef](#)] [[PubMed](#)]
15. Killian, M.L. Avian influenza virus sample types, collection, and handling. In *Animal Influenza Virus*; Humana Press: Totowa, NJ, USA, 2014; pp. 83–91.
16. Murphy, F.A.; Gibbs EP, J.; Horzinek, M.C.; Studdert, M.J. *Veterinary virology, Orthomyxoviridae*; Academic Press Inc.: San Diego, CA, USA, 1987.
17. Compans, R.W.; Oldstone, M.B. *Influenza Pathogenesis and Control-Volume I*; Springer: Berlin/Heidelberg, Germany, 2014.
18. Boyce, W.M.; Sandrock, C.; Kreuder-Johnson, C.; Kelly, T.; Cardona, C. Avian influenza viruses in wild birds: A moving target. *Comparative Immunol. Microbiol. Infect. Dis.* **2009**, *32*, 275–286. [[CrossRef](#)] [[PubMed](#)]
19. Jeong, J.; Woo, C.; Ip, H.; An, I.; Kim, Y.; Lee, K.; Jo, S.-D.; Son, K.; Lee, S.; Oem, J.-K.; et al. Identification of two novel reassortant avian influenza A (H5N6) viruses in whooper swans in Korea, 2016. *Virol. J.* **2017**, *14*, 60. [[CrossRef](#)] [[PubMed](#)]
20. Jiang, H.; Wu, P.; Uyeki, T.M.; He, J.; Deng, Z.; Xu, W.; Lv, Q.; Zhang, J.; Wu, Y.; Tsang, T.K.L.; et al. Preliminary epidemiologic assessment of human infections with highly pathogenic avian influenza A (H5N6) virus, China. *Clin. Infect. Dis.* **2017**, *65*, 383–388. [[CrossRef](#)] [[PubMed](#)]
21. Yoo, D.S.; Chun, B.; Min, K.D.; Lim, J.S.; Moon, O.K.; Lee, K.N. Elucidating the Local Transmission Dynamics of Highly Pathogenic Avian Influenza H5N6 in the Republic of Korea by Integrating Phylogenetic Information. *Pathogens* **2021**, *10*, 691. [[CrossRef](#)]
22. Lee, J.; Ko, Y.; Jung, E. Effective control measures considering spatial heterogeneity to mitigate the 2016–2017 avian influenza epidemic in the Republic of Korea. *PLoS ONE* **2019**, *14*, e0218202. [[CrossRef](#)]
23. Kim, H.K.; Jeong, D.G.; Yoon, S.-W. Recent outbreaks of highly pathogenic avian influenza viruses in South Korea. *Clin. Exp. Vaccine Res.* **2017**, *6*, 95–103. [[CrossRef](#)]
24. Qi, X.; Cui, L.; Yu, H.; Ge, Y.; Tang, F. Whole-genome sequence of a reassortant H5N6 avian influenza virus isolated from a live poultry market in China, 2013. *Genome Announc.* **2014**, *2*, e00706–e00714. [[CrossRef](#)]
25. Lee, E.-K.; Lee, Y.-N.; Kye, S.-J.; Lewis, N.S.; Brown, I.H.; Sagong, M.; Heo, G.-B.; Kang, Y.-M.; Cho, H.-K.; Kang, H.-M.; et al. Characterization of a novel reassortant H5N6 highly pathogenic avian influenza virus clade 2.3. 4.4 in Korea, 2017. *Emerg. Microbes Infect.* **2018**, *7*, 103. [[CrossRef](#)]

26. Lee, C.-W.; Suarez, D.L.; Tumpey, T.M.; Sung, H.-W.; Kwon, Y.-K.; Lee, Y.-J.; Choi, J.-G.; Joh, S.-J.; Kim, M.-C.; Lee, E.-K.; et al. Characterization of highly pathogenic H5N1 avian influenza A viruses isolated from South Korea. *J. Virol.* **2005**, *79*, 3692–3702. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Yoon, H.; Park, C.-K.; Nam, H.-M.; Wee, S.-H. Virus spread pattern within infected chicken farms using regression model: The 2003–2004 HPAI epidemic in the Republic of Korea. *J. Veter. Med. Ser. B* **2005**, *52*, 428–431. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Shin, J.; Kang, S.; Byeon, H.; Cho, S.M.; Kim, S.Y.; Chung, Y.J.; Jung, S.H. Highly pathogenic H5N6 avian influenza virus subtype clade 2.3. 4.4 indigenous in South Korea. *Sci. Rep.* **2020**, *10*, 7241. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Yoo, D.-S.; Chun, B.C.; Kim, Y.; Lee, K.-N.; Moon, O.-K. Dynamics of inter-farm transmission of highly pathogenic avian influenza H5N6 integrating vehicle movements and phylogenetic information. *Sci. Rep.* **2021**, *11*, 24163. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Ali, Z.; Hasan, M.; Giasuddin, M. Potential risk factors of avian influenza virus infection in asymptomatic commercial chicken flocks in selected areas of Bangladesh during 2019. *J. Adv. Veter. Animal Res.* **2021**, *8*, 51. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Murtagh, F.; Contreras, P. Methods of hierarchical clustering. *arXiv* **2011**, arXiv:1105.0121.
32. Kimes, P.; Liu, Y.; Hayes, D.N.; Marron, J.S. Statistical significance for hierarchical clustering. *Biometrics* **2017**, *73*, 811–821. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Ihaka, R.; Gentleman, R. R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.* **1996**, *5*, 299. [\[CrossRef\]](#)
34. Maugeri, A.; Barchitta, M.; Basile, G.; Agodi, A. Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across Italian regions. *Sci. Rep.* **2021**, *11*, 7082. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Husson, F.; Josse, J.; Pages, J. Principal component methods-hierarchical clustering-partitional clustering: Why would we need to choose for visualizing data. *Appl. Math. Dep.* **2010**, 1–17.
36. Gardener, M. *Beginning R: The Statistical Programming Language*; John Wiley & Sons: New York, NY, USA, 2012.
37. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [\[CrossRef\]](#)
38. Kassambara, A.; Mundt, F. Factoextra: Extract and visualize the results of multivariate data analyses. *R Package Version* **2017**, *1*, 337–354.
39. Lever, J.; Krzywinski, M.; Altman, N. Points of significance: Principal component analysis. *Nat. Methods* **2017**, *14*, 641–643. [\[CrossRef\]](#)
40. Szekely, J.G.; Rizzo, M.L. Hierarchical clustering via joint between-within distances: Extending Ward’s minimum variance method. *J. Classif.* **2005**, *22*, 151–184. [\[CrossRef\]](#)
41. Zolfaghari, F.; Khosravi, H.; Shahriyari, A.; Jabbari, M.; Abolhasani, A. Hierarchical cluster analysis to identify the homogeneous desertification management units. *PLoS ONE* **2019**, *14*, e0226355. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chem. Intell. Lab. Syst.* **1987**, *2*, 37–52. [\[CrossRef\]](#)
43. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [\[CrossRef\]](#)
44. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev.* **2010**, *2*, 433–459. [\[CrossRef\]](#)
45. Erişoğlu, M.; Sakallıoğlu, S. An investigation of effects on hierarchical clustering of distance measurements. *Selcuk. J. Appl. Math.* **2010**, 39–53.
46. Hill, S.C.; Lee, Y.-J.; Song, B.-M.; Kang, H.-M.; Lee, E.-K.; Hanna, A.; Gilbert, M.; Brown, I.H.; Pybus, O.G. Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea. *Infect. Genet. Evol.* **2015**, *34*, 267–277. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Wild Waterfowl Migration and Domestic Duck Density Shape the Epidemiology of Highly Pathogenic H5N8 Influenza in the Republic of Korea—Scientific Figure on ResearchGate. Available online: [https://www.researchgate.net/figure/Maps-showing-domestic-poultry-density-number-per-kilometer-colors-in-key-in-ROK\\_fig3\\_278412772](https://www.researchgate.net/figure/Maps-showing-domestic-poultry-density-number-per-kilometer-colors-in-key-in-ROK_fig3_278412772) (accessed on 24 February 2022).
48. Horby, P.; Sudoyo, H.; Viprakasit, V.; Fox, A.; Thai, P.Q.; Yu, H.; Davila, S.; Hibberd, M.; Dunstan, S.J.; Monteerarat, Y.; et al. What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1? *Epidemiol. Infect.* **2010**, *138*, 1550–1558. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Van Boven, M.; Koopmans, M.; Du Ry van Beest Holle, M.; Meijer, A.; Klinkenberg, D.; Donnelly, C.A.; Heesterbeek, H. Detecting emerging transmissibility of avian influenza virus in human households. *PLoS Comput. Biol.* **2007**, *3*, e145. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Pearl, D.L.; Louie, M.; Chui, L.; Dore, K.; Grimsrud, K.M.; Martin, S.W.; Michel, P.; Svenson, L.W.; McEwen, S.A. Epidemiological characteristics of reported sporadic and outbreak cases of *E. coli* O157 in people from Alberta, Canada (2000–2002): Methodological challenges of comparing clustered to unclustered data. *Epidemiol. Infect.* **2008**, *136*, 483–491. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Bouma, A.; Claassen, I.; Natih, K.; Klinkenberg, D.; Donnelly, C.; Koch, G.; Van Boven, M. Estimation of transmission parameters of H5N1 avian influenza virus in chickens. *PLoS Pathog.* **2009**, *5*, e1000281. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Tiensin, T.; Ahmed, S.S.U.; Rojanasthien, S.; Songserm, T.; Ratanakorn, P.; Chaichoun, K.; Kalpravidh, W.; Wongkasemjit, S.; Patchimasiri, T.; Chanachai, K.; et al. Ecologic risk factor investigation of clusters of avian influenza A (H5N1) virus infection in Thailand. *J. Infect. Dis.* **2009**, *199*, 1735–1743. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Wang, X.-X.; Cheng, W.; Yu, Z.; Liu, S.-L.; Mao, H.-Y.; Chen, E.-F. Risk factors for avian influenza virus in backyard poultry flocks and environments in Zhejiang Province, China: A cross-sectional study. *Infect. Dis. Poverty* **2018**, *7*, 65. [\[CrossRef\]](#)

- 
54. Martin, V.; Pfeiffer, D.U.; Zhou, X.; Xiao, X.; Prosser, D.J.; Guo, F.; Gilbert, M. Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. *PLoS Pathog.* **2011**, *7*, e1001308. [[CrossRef](#)]
  55. Velkers, F.C.; Manders, T.T.M.; Vernooij, J.C.M.; Stahl, J.; Slaterus, R.; Stegeman, J.A. Association of wild bird densities around poultry farms with the risk of highly pathogenic avian influenza virus subtype H5N8 outbreaks in the Netherlands, 2016. *Transbound. Emerg. Dis.* **2021**, *68*, 76–87. [[CrossRef](#)]