Multiple Kernel Learning with Maximum Inundation Extent from MODIS Imagery for Spatial Prediction of Flood Susceptibility

Qiang Hu¹ · Yuelong Zhu¹ · Hexuan Hu¹,² · Zhuang Guan¹ · Zeyu Qian¹ · Aiming Yang¹,²,³,⁴,⁵

Received: 5 July 2021 / Accepted: 21 October 2021 / Published online: 8 November 2021 © The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
Identifying flood prone areas is essential for basin management. In this paper, a spatial prediction technology of flood susceptibility based on multiple kernel learning (MKL) is proposed. We establish the flood susceptibility model by using EasyMKL, nonlinear MKL (NLMKL), Representative MKL (RMKL), Generalized MKL (GMKL), support vector machine (SVM) with linear kernel and SVM with Gaussian radial base function (RBF) kernel. The spatial prediction of flood susceptibility in Sanhuajian basin of the Yellow River is carried out. We use MODIS remote sensing images to obtain historical flood inundation sites in the study area. Then, ten flood conditioning factors are used as inputs to the flood susceptibility model. The model performance is evaluated in terms of accuracy (ACC), balanced F Score (F1 score), and areas under the curve (AUC). According to the results, MKL significantly outperforms the SVM adopting single kernel, and NLMKL (ACC = 0.833, F1 = 0.841, AUC = 0.889) demonstrates the best comprehensive performance. The flood susceptibility map generated by MODIS remote sensing images and MKL, therefore, can provide effective help for researchers and decision makers in flood management.

Keywords Flood susceptibility · Multiple kernel learning · Support vector machine · Remote sensing
1 Introduction

Flood is one of the most destructive natural disasters. It inflicts massive losses all over the world on a yearly basis (Chen et al. 2020). From 1980 to 2016, flood accounted for 23% of global economic loss (Mateo et al. 2017). Such losses can be mitigated or even averted by effective flood risk management methods.

Hydrodynamics-based flood inundation models, such as HEC-RAS (Pappenberger et al. 2005), Mike21 (Beilicci et al. 2016), Delft3D (Hoch et al. 2018), LISFLOOD-FP (Rajib et al. 2020), etc., can provide high-precision information on the spatial distribution of flood risks in a river basin so as to facilitate decision-making. However, it is not easy to use these models. These models require researchers to have high professional knowledge, and model calculation requires detailed basin data, a large amount of computing resources, and each calculation case takes a long time (Teng et al. 2017). Therefore, they are not suitable for basin scale flood assessment.

In recent years, flood susceptibility map (FSM) combined with machine learning and geographical information system (GIS) technology has been widely used in flood management (Tehrany et al. 2013; Shafapour et al. 2019; Zhao et al. 2019; Liuzzo et al. 2019). Machine learning model of FSM identifies flood prone areas by learning the relationship between historical flood records and flood conditioning factors. At present, the commonly used methods include SVM (Tehrany et al. 2014; Zhao et al. 2019), artificial neural network (Kia et al. 2012; Tien et al. 2020), decision tree (Tehrany et al. 2013; Yariyan et al. 2020), Bayesian method (Janizadeh et al. 2021).

MKL is a machine learning method based on SVM. MKL, now, has become a hot-spot in machine learning researches and been widely used in imaging (Gu et al. 2016), bioinformatics (Qi et al. 2020), and finance (Arratia et al. 2020), but it has not been applied to the research of FSM. MKL learns through a combination of multiple kernel functions instead of one, and the kernel function selection and parameter optimization are the key to performance. The kernel combination can be linear, nonlinear, or data dependent (Gönen and Alpaydin 2011). Compared with SVM using single kernel function, MKL method has better generalization ability and robustness.

In the actual modelling process, however, insufficiency and poor availability of the historical data (labeled data) is a commonplace issue. Moreover, since the majority of studies only focus on a single flood outbreak in the study area, the findings are prone to have limitations and do not consider the comprehensive situation of multiple floods in the basin over a long period of time. In view of this, we employed the open water likelihood index (OWL) method to extract the inundation range from MODIS remote sensing images in the study area (Huang et al. 2014). The historical maximum flood inundation range in the study area is obtained by combining the inundation ranges in different periods, and used this as the base map for the sampling of the machine learning model.

This research focuses on using different combinations of MKL to classify the flood susceptibility in the study area and draw the FSM. The study area is from Sanmenxia to Huayuankou of the Yellow River Basin in China. EasyMKL, a scalable MKL algorithm (Aiolli and Donini 2015) and NLMKL (Cortes et al. 2009) are adopted, compared with SVM (linear kernel), SVM (RBF kernel), RMKL (Do et al. 2009), and GMKL (Varma and Babu 2009), and their performance is evaluated.
This paper examines the area between Sanmenxia and Huayuanko (A.K.A. the Sanhuajian Basin) in the middle reaches of the Yellow River, the longitude between 109.73° E and 113.58° E, and the latitude between 33.64° N and 36.97° N. The area of the watershed is 40,000km². Sanhuajian is surrounded by mountains in the north, west and south; the Yellow River runs from west to east across it in the middle. The two tributaries of Yiluo River and Qin River join the Yellow River above Huayuankou. The whole terrain of Sanhuajian is high on the north and south sides and low in the middle. Its altitude is about 2000 meters in the north and 1600 ~ 1800 meters in the south. In the middle, it gradually drops below 100 meters from the north and south sides. A semi-humid area, Sanhuajian witnesses frequent rainstorms from June to August, the annual precipitation is 600 ~ 800 mm, and the precipitation distribution decreases from south to north. (Wang and Zheng 2014).

3 Material and Methods

3.1 Historical Daily Flow Data

We obtained the historical daily flow data of the Huayuankou Site from the YRCC website¹, which cover ten complete water years (from 2007 to 2017) as shown in Fig. 1. Flooding is usually caused by flood peaks, and the major peaks are represented by red dots in the Fig. 1.

¹ http://www.yrcc.gov.cn/
3.2 MODIS Imagery Maximum Inundation Extent and Flood Inventories

MOD09A1 is a synthetic product of the MODIS surface reflectance by the Terra satellite following an exclusion of interferential factors like atmospheric radiation and aerosol. Its spatial resolution is 500 meters, and its temporal resolution is once every eight days (Barnes et al. 1998). MODIS remote sensing imagery is widely used in flood detection and suitable for large-scale spatial research (Chen et al. 2013; Coltin et al. 2016). This study adopts the open water likelihood (OWL) index (Huang et al. 2014) to extract flood inundation ranges from MOD09A1 remote sensing images. The OWL index contains five parameters sensitive to water bodies, namely, the short-wave infrared band (band 6 and band 7), NDVI (Townshend and Justice 1986), NDWI (Gao 1995) and the multi-resolution valley bottom flatness index (MrVBF) (Gallant and Dowling 2003). Modeled by nonlinear logistic regression, the OWL index adopts the expression below:

\[
\text{OWL} = \frac{1}{1 + \exp(f)}
\]

where

\[
f = a_0 + \sum_{i=1}^{5} a_i x_i
\]

where \(x_1 = \text{SWIR Band 6 (reflectance } \times 1000); x_2 = \text{SWIR Band 7 (reflectance } \times 1000); x_3 = \text{NDVI}; x_4 = \text{NDWI}; x_5 = \text{MrVBF}; a_0 = -3.41375620; a_1 = -0.000959735270; a_2 = 0.00417955330; a_3 = 14.1927990; a_4 = -0.430407140; a_5 = -0.0961932990.\) The OWL index value ranges from 0 to 100, representing the probability of water existence in each pixel of the MOD09A1 remote sensing images. Whether the pixel is a flooded or non-flooded one is determined by setting the corresponding threshold. In this study, the threshold is set as 0, so if the value exceeds 0, it is judged to be a flooded pixel.

The flood inundation range of a single image is obtained through the OWL index. ArcGIS can help synthesize multiple flood inundation ranges of the same study area throughout a long time series into the maximum flood inundation range, which is used as the base map for the sampling of the machine learning model. The chosen sample sites are labeled “flooded” or “non-flooded”.

According to the time points of flood peak occurrences in Fig. 1, we pinpointed the corresponding dates of the MOD09A1 remote sensing images, along with the dates before and after the occurrences. The flood inundation range of each remote sensing image was extracted by OWL index and synthesized via ArcGIS into the maximum flood inundation range of the study area.

In the maximum flood inundation range, 310 flooded sites were randomly selected; in the non-flood inundation range, 310 non-flooded sites were randomly selected. In Fig. 2, the red dots stand for flooded sites, and the blue dots, non-flooded sites. We used 70% (434) of the total sample sites as the training set and 30% (186) as the test set, marking the flooded sites as “1” and the non-flooded sites as “-1”.

3.3 Flood Conditioning Factors

Flood disasters involve an extremely complex process and can be attributed to many interplaying factors. Drawing reference from previous studies (Zhao et al. 2019; Dodangeh et al. 2020; Tien et al. 2020), this paper selects out ten flood conditioning factors, including elevation, slope, aspect, curvature, Stream Power Index (SPI), Topographic Wetness Index
(TWI), distance from river, soil type, vegetation, and annual average rainfall. Table 1 lists the data source, format, spatial and temporal resolution of these flood conditioning factors.

Topography mainly affects the occurrence and distribution of flood through elevation and topographic fluctuation, which is also considered to be the main factor affecting the occurrence and distribution of flood. The elevation data of the study area is ASTER GDEM V2 product, which is downloaded from "Geospatial Data Cloud"\(^2\). The slope, aspect, curvature and distance to rivers are obtained by processing DEM data with ArcGIS software.

**Table 1** Data source of flood conditioning factors

| name             | data format | spatial resolution | temporal resolution | source                                      |
|------------------|-------------|--------------------|---------------------|---------------------------------------------|
| DEM              | TIF         | 30m                | 2015                | ASTER GDEM V2                               |
| slope            | TIF         | 30m                | 2015                | DEM                                         |
| aspect           | TIF         | 30m                | 2015                | DEM                                         |
| curvature        | TIF         | 30m                | 2015                | DEM                                         |
| SPI              | TIF         | 30m                | 2015                | DEM                                         |
| TWI              | TIF         | 30m                | 2015                | DEM                                         |
| Distance from river | shapefile |                   |                     | DEM                                         |
| Soil types       | shapefile   | 1:1000000          | 1995                | CASEarth Data Sharing and Service Portal    |
| vegetation       | shapefile   | 1:1000000          | 2007                | CASEarth Data Sharing and Service Portal    |
| rainfall         | TIF         | 0.5° × 0.5°        | 2007-2017           | 0.5° × 0.5° raster data set of monthly surface precipitation in China (V2.0) |

\(^2\) [http://www.gscloud.cn](http://www.gscloud.cn)
Stream Power Index (SPI) represents the river erosivity of the catchment area, while Topographic Wetness Index (TWI) takes into account the influence of the terrain and soil characteristics on soil moisture distribution (Regmi et al. 2010). They are calculated using the following formulas:

\[ SPI = A \tan \beta / b \]  

\[ TWI = \ln(A/b \tan \beta) \]

where \( A \) (m\(^2\)) is the upstream catchment area or flow accumulation, \( b \) (m) is the unit width of water flow, and \( \beta \) (radian) is the slope.

Vegetation type and soil type jointly determine water infiltration, water holding and water control capacity. The data of soil type and vegetation type comes from "earth big data engineering data sharing service system"\(^3\).

The data of annual average rainfall for many years are from "Daily value of surface precipitation in China" in 0.5° × 0.5° grid data set (V2.0) \(^4\) in " China Meteorological Data Service Centre", we selected the data of the study area from 2007 to 2017, calculated the annual average rainfall of 10 years by using ArcGIS, and formed the multi-year average rainfall distribution map by spline function interpolation.

Because the resolution of each factor is different, we used ArcGIS to preprocess all the data and resampled them into raster data with a resolution of 90m, so that all raster data share the same coordinate system and pixel size, making it convenient for machine learning and statistical analysis. Flood conditioning factors are shown in Fig. 3.

**3.4 Multiple Kernel Learning**

Support vector machines often use a single kernel function to measure the similarity between different instances. The selection of the kernel function and its parameters determines the structure of feature space as well as the classification effect (Damasevicius 2010). However, a single kernel function cannot fully describe data similarity. MKL, in contrast, combines multiple kernel functions and thus often outperforms single kernel SVM. The kernel combination of MKL can be linear (weighted addition of multiple kernels), nonlinear (product or power of multiple kernels) or data dependent (kernel combination in relation to sample characteristics). By finding the best kernel combination parameters, MKL combines several base kernels and makes use of their respective advantages to obtain a kernel function with stronger learning and generalization abilities. Common methods of determining the kernel combination parameters include the fixed rule-based method, heuristic method, and optimization method (Gönen and Alpaydin 2011). This paper introduces EasyMKL of linear kernel combination and NLMKL of nonlinear kernel combination. Both of them use optimization method to determine the parameters of kernel combination.

**3.4.1 EasyMKL**

EasyMKL (Aiolli and Donini 2015), a highly scalable MKL algorithm, can effectively handle hundreds of thousands of kernels. Compared with other MKL algorithms (e.g. RMKL, GMKL, etc.), it shows better performance and maintains good robustness

---

\(^3\) [http://data.casearth.cn/](http://data.casearth.cn/)

\(^4\) [http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_PRE_MON_GRID_0.5.html](http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_PRE_MON_GRID_0.5.html)
even in the presence of data noise (Aiolli and Donini 2015; Saleh et al. 2018; Arratia et al. 2020). This algorithm has a linear combination of the base kernels in the following form:
\[ K = \sum_{r=1}^{R} \eta_r K_r, \eta_r \geq 0 \]  

where \( K_r \) is the base kernel function, \( R \) is the number of base kernels, and \( \eta_r \), the weight coefficient of each base kernel. EasyMKL obtains the best combination parameters by solving a min-max problem about the variables \( \gamma \) and \( \eta \):

\[
\max_{\eta, \gamma} \min_{r \in \mathbb{R}} (1 - \lambda)\gamma^T \hat{Y} \left( \sum_{r}^{R} \eta_r \hat{K}_r \right) \hat{Y} \gamma + \lambda \| \gamma \|^2
\]

where \( \gamma \) is the probability distribution of positive and negative training samples, \( \eta \) is the base kernel combination vector, a unitary norm vector. \( \hat{Y} \) is the label vector of training set, and \( \hat{K}_r \), the \( r \)-th base kernel. Define the vector of the \( r \)-th entry \( d_r(\gamma) = \gamma^T \hat{Y} \hat{K}_r \hat{Y} \gamma \), and then the optimization problem can be written as:

\[
\min_{r \in \mathbb{R}} \max_{\eta \in \mathbb{R}} (1 - \lambda)\eta^T d(\gamma) + \lambda \| \gamma \|^2
\]

By maximizing the distance between positive and negative samples, the min-max problem can be simplified as a quadratic problem, of which the analytical solution is as follows:

\[
\eta^* = \frac{d(\gamma^*)}{\|d(\gamma^*)\|^2}
\]

The weight parameters of each base kernel \( K_r \) are obtained through the following formula:

\[
\eta_r = \gamma^* \hat{Y} \left( \hat{K}_r / \hat{T}_r (\hat{K}_r) \right) Y \gamma^*
\]

### 3.4.2 NLMKL

NLMKL is a nonlinear polynomial kernel combination method (Cortes et al. 2009):

\[
K = \sum_{k_1, \ldots, k_d \geq 0} \eta_{k_1}^{k_1} \cdots \eta_{r}^{k_r} K_1^{k_1} \cdots K_r^{k_r}, \eta_r \geq 0
\]

where \( d \) is the degree of the combination of polynomials, \( K \) is a positive semidefinite matrix. In order to reduce the complexity, only the case where \( d = 2 \) is taken into consideration, and thus the nonlinear kernel combination can be given the following expression:

\[
K = \sum_{k=1}^{R} \sum_{l=1}^{R} \eta_k \eta_l K_k K_l
\]

Apply (11) to SVM, and then the weight coefficient \( \eta_r \) of each base kernel is obtained by solving a min-max problem:

\[
\min_{\eta \in \mathbb{R}} \max_{\alpha \in \mathbb{R}^m} \sum_{i=1}^{m} \alpha_i \cdot \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K
\]
where $\Omega$ is a positive, bounded, and convex set. Norm-1 and norm-2 bounded sets constitute the two possible choices for the set $\Omega$, and the norm-2 bounded set is adopted in this paper.

$$\Omega_2 = \{ \eta \mid \eta \geq 0 \land \|\eta - \eta_0\|_2 \leq \Lambda \}$$ (13)

In Eq. (13), $\eta_0$ is the bias parameter of the base kernel weight vector $\eta$. In actual implementation, $\eta_0$ is initially taken as 0 and $\Lambda$ as 1. The min-max optimization problem of Eq. (8) can be solved by a two-step method. This paper proposes a momentum-based gradient descent algorithm (Algorithm 1). In each iteration, firstly, given the fixation of the base kernel weight vector $\eta$, the support vector machine problem is solved to obtain $\alpha$. Then, considering the constraint $\Omega_2$, the base kernel weight vector $\eta$ is updated by the gradient descent method with momentum term while the SVM parameters are fixed. Repeat these two steps until convergence. Because the momentum term increases gradually in the gradient direction while decreasing gradually where the gradient direction changes, we can get faster convergence speed and weaker oscillation.

As the focus of this paper is to develop a machine learning model for spatial prediction of flood susceptibility, we have compared the two proposed MKL models with other algorithms, selecting SVM (linear kernel), SVM (RBF kernel), RMKL, and GMKL as baseline methods. The setting particulars of the four MKL algorithms (EasyMKL, NLMK, RMKL and GMKL), all of which use linear and RBF kernels as the base kernel, will be covered in Sect. 3.5.

---

**Algorithm 1 Momentum-Based Gradient Descent**

**Input:** $\eta_{\text{init}} \in \Omega_2, \delta \in [0,1], \beta \in [0,1] : \epsilon > 0, (K_k)_{k=1}^p, \eta' \leftarrow \eta_{\text{init}}, v_t \leftarrow v_{\text{init}}$  

**Repeat:**

$$\eta \leftarrow \eta', K = \sum_{k=1}^p \sum_{t=1}^T \eta_k \eta_t K_k K_t$$

$$\alpha \leftarrow \text{SVM}_{\text{step}}$$

$$v_t \leftarrow \delta \alpha^T K \alpha + \beta v_t$$

$$\eta' \leftarrow -v_t + \eta$$

$$\forall k, \eta'_k \leftarrow \max \left(0, \eta'_k\right)$$

**normalize** $\eta'$, s.t. $\|\eta' - \eta_0\| = \Lambda$

**Until:** $\|\eta' - \eta\| < \epsilon$

---

### 3.5 Experiment Settings

The historical flood inundation data set (Sect. 3.2) includes 620 instances (310 flooded sites and 310 non-flooded sites). Randomly select 70% (434) of the instances as the training set and 30% (186) of the instances as the test set. The number of flooded sites and non-flooded sites in training set is equal, and so is the test set. The dimension (feature length) of each instance is 10. In order to get the best performance, the data set has been standardized.

SVM (linear kernel), SVM (RBF kernel), RMKL and GMKL were adopted as the baseline methods. EasyMKL, NLMK, RMKL and GMKL all used linear and RBF kernels as the base kernels.
We implemented these algorithms in Python using Sklearn Library and PyTorch Library and found the best hyper parameter configuration through 5-fold cross-validation and grid search: C for SVM and γ for RBF kernel function. The value range of C is $C = \{ 10^{-2}, 10^{-1}, \ldots, 10^{2} \}$, and that of γ is $\gamma = \{0.01, 0.1, 0.2, 0.5, 0.8, 1\}$. Accuracy is the evaluation criterion of the cross-validation. The table of optimal parameter selection is shown in Appendix A.

### 3.6 Performance Evaluation

The performance of each model was evaluated in terms of accuracy (ACC), F1 score, and Area Under Curve (AUC) – common indicators for machine learning model evaluation (Choubin et al. 2019; Hu et al. 2021). TP, TN, FP, and FN were true positive (correctly classified flooded sites), true negative (correctly classified non-flooded sites), false positive (misclassified flooded sites) and false negative (misclassified non-flooded sites), respectively.

\[
ACC = \frac{TP + TN}{TP + FP + TN + FN}
\]

\[
F1 = \frac{2 \times P \times R}{P + R}
\]

\[
P = \frac{TP}{TP + FP}
\]

\[
R = \frac{TP}{TP + FN}
\]

As to the receiver operating characteristic (ROC) curve, false positive rate (FPR) is defined as the X-axis, and true positive rate (TPR), the Y-axis. Area Under Curve (AUC), which is widely applied in FSM model evaluation, was used to assess the model performance. When AUC is 0.5, it means the model works like a random model; when AUC exceeds 0.8, it shows the model has good performance (Tien et al. 2016); when AUC is 1, the model features perfect classification.

### 4 Experimental Results and Discussion

#### 4.1 Experiment Results and Analysis

Table 2 presents a detailed comparison of the classification algorithms used in the training set. The performance of the multiple kernel method is better than that of the single

|           | SVM(linear) | SVM(RBF) | EasyMKL | NLMKL | RMKL | GMKL |
|-----------|-------------|----------|---------|-------|------|------|
| ACC       | 0.800       | 0.876    | 0.917   | 0.938 | 0.901| 0.912|
| F1        | 0.811       | 0.884    | 0.92    | 0.941 | 0.904| 0.916|
| P         | 0.786       | 0.851    | 0.908   | 0.919 | 0.878| 0.9  |
| R         | 0.838       | 0.919    | 0.933   | 0.964 | 0.933| 0.933|
In terms of accuracy, NLMKL is the best (0.938), followed by EasyMKL (0.917) and GMKL (0.912). In terms of F score, NLMKL (0.941) and EasyMKL (0.92) are the best performers. Figure 4 shows the ROC curve and AUC value of each model in the training set. EasyMKL (0.973) and NLMKL (0.987) have higher AUC values, similar ROC curves, and therefore similar performances, followed by GMKL (0.970).

Table 3 presents a detailed comparison of the classification algorithms used in the test set. In general, the multiple kernel method again outperforms the single kernel support vector machine. In terms of accuracy, NLMKL is the best (0.833), followed by EasyMKL (0.823) and GMKL (0.817). With regard to the F score, the ranking remains unchanged. Figure 5 shows the ROC curve and AUC value of each model in the training set. All classification algorithms show similar performance, yet NLMKL (0.889) has a higher AUC value.

The trained machine learning model was used to predict the probability of flood inundation for each pixel in the study area. The classification methods commonly used in GIS include natural breaks, quantile, equal interval, and geometric interval (Tehrany et al. 2013). The natural breaks method, based on the inherent natural grouping of data, can identify the similar values and group them in the most appropriate manner (Smith et al. 2007). It can also maximize the difference between classes and set the boundary at the point where the difference of data values is relatively large. Therefore, we used the natural breaks method in ArcGIS to divide the model-obtained probability values into five categories (very low, low, moderate, high, and very high) and completed the flood susceptibility map of the study area (Fig. 6). As can be seen from all the figures, the most flood-vulnerable areas are mainly located in the east of the basin. These areas are low and flat, and placed in proximity of the river and most of the vegetation there is cultivated.

**Table 3** Performance of different models in test set

|          | SVM(linear) | SVM(RBF) | EasyMKL | NLMKL | RMKL | GMKL |
|----------|-------------|----------|---------|-------|------|------|
| ACC      | 0.806       | 0.801    | 0.823   | 0.833 | 0.812| 0.817|
| F1       | 0.816       | 0.810    | 0.833   | 0.841 | 0.821| 0.821|
| P        | 0.734       | 0.731    | 0.762   | 0.759 | 0.741| 0.757|
| R        | 0.919       | 0.908    | 0.92    | 0.943 | 0.919| 0.897|
Flood density was taken into consideration to further evaluate the performance of the flood susceptibility map (Chapi et al. 2017). FD is the ratio of flooded site number to class pixel number in each susceptibility category. Ideally, density values should increase from very low susceptibility areas to very high susceptibility areas (Pradhan and Lee 2010). Table 4 lists the distribution of the categories of flood susceptibility maps by each model, the number of flood sites in each category, and flood density. Figure 7 shows flood density in different flood susceptibility areas by the six compared models. The results show that the flood susceptibility map obtained by the NLMKL (363.68) model has the highest FD in extremely high susceptibility area. Also a multiple kernel learning method, EasyMKL
Multiple Kernel Learning with Maximum Inundation Extent from…

(360.66) shows great performance, which is superior to other baseline methods. Likewise, in very low and low susceptibility areas, the multiple kernel learning method outperforms the single kernel SVM. Among those in comparison, NLMKL (0.72 and 2.87) is the best performer. In terms of flood site coverage, Fig. 8 shows the inundation sites in different flood susceptibility areas by the six compared models. EasyMKL and NLMKL cover 283 and 287 flood sites respectively in the high and very high susceptibility areas, outperforming the linear kernel SVM (254), RBF kernel SVM (269), RMKL (271), and GMKL (275); in low and very low susceptibility areas, EasyMKL and NLMKL cover six and four flood sites respectively, outperforming the linear kernel SVM (11), RBF kernel SVM (16), RMKL (11), and GMKL (10).
Discussion

FSM can help managers and decision-makers identify flood susceptibility areas in the basin. The establishment of FSM model requires historical inundation site data in the study area, but the data is valuable and difficult to obtain. We found that MODIS remote sensing can be used to obtain historical inundation site data, and the six models used in this paper have achieved good performance.

![Flood density in different flood susceptibility areas by the six models](image1)

![Flooded sites in different flood susceptibility areas by the six models](image2)

**Fig. 7** Flood density in different flood susceptibility areas by the six models

**Fig. 8** Flooded sites in different flood susceptibility areas by the six models
FSM generated by the six models are relatively similar. In the study area of this paper, there is a strong correlation between flood susceptibility and river network. The distribution of river network determines the possibility of flood inundation in a certain place to a certain extent. The closer to the river, the higher susceptibility of flood. The susceptibility of flood is highest in plain areas close to the watershed outlet. Topography is also the main factor affecting flood occurrence and distribution. Areas with low altitude and gentle slope have high flood susceptibility. During the flood season, these areas need key protection.

Among the six models tested in this paper, the flood susceptibility prediction using NLMKL has the highest accuracy and performance, and the MKL performance is better than that of single kernel SVM. Central to MKL is the combination of multiple kernel functions and the learning method for the base kernel weight coefficient. For NLMKL with nonlinear kernel function combination, a momentum-based gradient descent algorithm is proposed to solve the optimal weight of each kernel matrix. This method can eliminate the heterogeneous information in sample features and uneven distribution of data in high-dimensional feature space, and quickly and stably obtain the optimal parameters and optimal model.

6 Conclusion

In this study, two multiple kernel learning methods were developed for flood susceptibility mapping of Sanhuajian of the Yellow River. The performance of the two MKL methods (EasyMKL and NLMKL) is compared with that of the linear kernel SVM, RBF kernel SVM, RMKL, and GMKL. The results show that the MKL method has better performance and is more suitable for flood susceptibility mapping, especially for flood inundation sites extracted from historical MODIS remote sensing images alone. To summarize, the major contributions of this study are as follows:

1. In the absence of any historical records of the inundation sites, the sites are pinpointed using MODIS remote sensing images to provide data sets for machine learning models and complete the flood susceptibility map of the study area. This data acquisition method can provide reference to other areas in need of a flood susceptibility map when lacking a record of the inundation sites.
2. The two MKL models have higher prediction accuracy and higher quality in generating the flood susceptibility map. Their performance is better than that of the single kernel support vector machine.
3. Between the two MKL methods, NLMKL has the higher prediction accuracy, F1 score, and AUC than EasyMKL. In terms of the flood susceptibility map performance, NLMKL is also slightly better than EasyMKL.

Appendix A

Table 5 shows the best parameter configuration of the implemented algorithms in paper.
Appendix B

Figure 9 shows the maximum flood inundation map, which is extracted from MODIS imagery using OWL algorithm and ArcGIS.

Table 5 Best parameter configuration

| Method      | C   | $\gamma$ | Accuracy |
|-------------|-----|----------|----------|
| SVM(linear) | 1   | -        | 0.800    |
| SVM(RBF)    | 1   | 0.1      | 0.876    |
| EasyMKL     | 1   | 0.1      | 0.917    |
| NLMKL       | 1   | 0.2      | 0.938    |
| RMKL        | 1   | 0.1      | 0.901    |
| GMKL        | 0.1 | 0.1      | 0.912    |

Fig. 9 maximum flood inundation map
Multiple Kernel Learning with Maximum Inundation Extent from…

Authors Contributions Q. Hu: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Roles/Writing - original draft. Y.L. Zhu: Funding acquisition, Project administration, Writing - review & editing. H.X. Hu: Conceptualization, Methodology, Supervision, Writing - review & editing. Z. Guan: Data curation, Software, Visualization. Z.Y. Qian: Data curation, Software, Visualization. A. Yang: Resources, Project administration.

Funding This work was supported by National Key Research and Development Program of China (Grant No: 2018YFC0407904).

Declarations

Ethical Approval All authors have seen and agreed with the contents of the manuscript and are looking forward to publishing this paper on “Water Resources Management” journal.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Conflict of Interest The authors declare that there is no conflict of interest.

Competing Interests The authors declare no competing interests.

References

Aiolli F, Donini M (2015) EasyMKL: a scalable multiple kernel learning algorithm. Neurocomputing 169:215–224. https://doi.org/10.1016/j.neucom.2014.11.078

Arratia A, Belanche LA, Fábregues L (2020) An evaluation of equity premium prediction using multiple kernel learning with financial features. Neural Process Lett 52(1):117–134. https://doi.org/10.1007/s11063-018-09971-7

Barnes WL, Pagano TS, Salomonson VV (1998) Prelaunch characteristics of the moderate resolution imaging spectroradiometer (modis) on eos-am1. IEEE Trans Geosci Remote Sens 36(4):1088–1100. https://doi.org/10.1109/36.700993

Beilicci E, Beilicci R, Visescu M (2016) Study of flooding risk areas using 1D/2D advanced hydroinformatic tools. Int Multidiscip Sci GeoConf SGEM

Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. Modell Softw 95:229–245. https://doi.org/10.1016/j.envsoft.2017.06.012

Chen B, Shi F, Lin T, Shi P, Zheng J (2020) Intensive Versus Extensive Events? Insights from Cumulative Flood-Induced Mortality Over the Globe, 1976–2016. Int J Disaster Risk Sci 11:441–451. https://doi.org/10.1007/s13753-020-00288-5

Chen Y, Huang C, Ticehurst C, Merrin L, Thew P (2013) An evaluation of MODIS daily and 8-day composite products for floodplain and wetland inundation mapping. Wetlands 33(5):823–835. https://doi.org/10.1007/s11757-013-0439-4

Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci Total Environ 651:2087–2096. https://doi.org/10.1016/j.scitotenv.2018.10.064

Coltin B, McMichael S, Smith T et al (2016) Automatic boosted flood mapping from satellite data. Int J Remote Sens 37(5):993–1015. https://doi.org/10.1080/01431161.2016.1145366

Cortes C, Mohri M, Rostamizadeh A (2009) Learning non-linear combinations of kernels. Adv Neural Inf Process Syst 396–404

Damasevicius R (2010) Optimization of SVM parameters for recognition of regulatory DNA sequences. Top 18(2):339–353. https://doi.org/10.1007/s11750-010-0152-x

Dodangeh E, Choubin B, Eigdir AN, Nabipour N, Panahi M, Shamshirband S, Mosavi A (2020) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. Sci Total Environ 705:135983. https://doi.org/10.1016/j.scitotenv.2019.135983
Do H, Kalousis A, Woznica A, Hilario M (2009) Margin and radius based multiple kernel learning. Machine Learning and Knowledge Discovery in Databases, Springer, Berlin, Heidelberg, pp 330–343

Gao BC (1995) NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ 58:257–266. https://doi.org/10.10117/12.210877

Gallant JC, Dowling TI (2003) A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour Res 39(12):1347–1359. https://doi.org/10.1029/2002WR001426

Gönen M, Alpaydin E (2011) Multiple kernel learning algorithms. J Mach Learn Res 12:2211–2268. https://doi.org/10.5555/1953048.2021071

Gu Y, Liu T, Jia X, Benediktsson JA, Chanussot J (2016) Nonlinear multiple kernel learning with multi-structure-element extended morphological profiles for hyperspectral image classification. IEEE Trans Geosci Remote Sens 54(6):3235–3247. https://doi.org/10.1109/TGRS.2015.2514161

Hoch JM, van Beek R, Winsemius HC, Bierkens MFP (2018) Benchmarking flexible meshes and regular grids for large-scale fluvial inundation modelling. Adv Water Resour 121:350–360. https://doi.org/10.1016/j.adwres.2018.09.003

Huang C, Chen Y, Wu J (2014) Mapping spatio-temporal flood inundation dynamics at large river basin scale using time-series flow data and MODIS imagery. Int J Appl Earth Obs Geoinf 26:350–362. https://doi.org/10.1016/j.jag.2013.09.002

Hu H, Li Q, Zhao Y, Zhang Y (2021) Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors. IEEE Trans Ind Inform 17(4):2880–2889. https://doi.org/10.1109/TII.2020.3022912

Janizadeh S, Vafakhah M, Kapelan Z et al (2021) Novel bayesian additive regression tree methodology for flood susceptibility modeling. Water Resour Manag. https://doi.org/10.1007/s11269-021-02972-7

Kia MB, Pirasteh S, Pradhan B et al (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ Earth Sci 67(1):251–264. https://doi.org/10.1007/s12665-011-1504-z

Liuzzo L, Sammartano F, Freni G (2019) Comparison between different distributed methods for flood susceptibility mapping. Water Resour Manag 33:3155–3173. https://doi.org/10.1007/s11269-019-02293-w

Mateo CMR, Yamazaki D, Kim H, Champathong A, Vaze J, Oki T (2017) Impacts of spatial resolution and representation of flow connectivity on large-scale simulation of floods. Hydrol Earth Syst Sci 21:5143–5163. https://doi.org/10.5194/hess-21-5143-2017

Pappenberger F, Beven K, Horritt M, Blazkova S (2005) Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. J Hydrol 302(1–4):46–69. https://doi.org/10.1016/j.jhydrol.2004.06.036

Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Modell Softw 25(6):747–759. https://doi.org/10.1016/j.envsoft.2009.10.016

Qi R, Wu J, Guo F, Xu L, Zou Q (2020) A spectral clustering with self-weighted multiple kernel learning method for single-cell RNA-seq data. Brief Bioinform 22(24):bbaa216. https://doi.org/10.1093/bib/bbaa216

Rajib A, Liu Z, Merwade V, Tavakoly A, Follum M (2020) Towards a large-scale locally relevant flood inundation modeling framework using SWAT and LISFLOOD-FP. J Hydrol 581:124406. https://doi.org/10.1016/j.jhydrol.2019.124406

Regmi NR, Giardino JR, Vitrek JD (2010) Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology 115(1–2):172–187. https://doi.org/10.1016/j.geomorph.2009.10.002

Saleh A, Abdel-Nasser M, Angel GM et al (2018) Aggregating the temporal coherent descriptors in videos using multiple learning kernel for action recognition. Pattern Recogn Lett 105:4–12. https://doi.org/10.1016/j.patrec.2017.06.010

Shafapour TM, Kumar L, Neamah JM, Shabani F (2019) Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. Geomatics Nat Hazards Risk 10(1):79–101. https://doi.org/10.1080/19475705.2018.1506509

Smith MJ, Goodchild MF, Longley PA (2007) Geospatial analysis: a comprehensive guide to principles, techniques and software tools (2. Ed.). DBLP

Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J Hydrol 404:69–79. https://doi.org/10.1016/j.jhydrol.2013.09.034

Tehrany MS, Pradhan B, Jebur MN (2014) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J Hydrol 512:332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008
Teng J, Jakeman AJ, Vaze J, Croke BFW, Dutta D, Kim S (2017) Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. Environ Modell Softw 90:201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Tien BD, Hoang N, Martínez-álvarez F, Ngo PT, Hoa PV, Pham TD, Samui P, Costache R (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. Sci Total Environ 701:134413. https://doi.org/10.1016/j.scitotenv.2019.134413

Tien BD, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13(2):361–378. https://doi.org/10.1007/s10346-015-0557-6

Townshend JRG, Justice CO (1986) Analysis of the dynamics of African vegetation using the normalized difference vegetation index. Int J Remote Sens 7:1435–1445. https://doi.org/10.1080/01431648608948946

Varma M, Babu B (2009) More generality in efficient multiple kernel learning. Int Conf Mach Learn ACM 1065–1072

Wang Z, Zheng S (2014) Application of the meteorological and hydrological models coupling in the flood forecasting in Sanmenxia-Huaiyuan reach of the Yellow River. Meteorol Environ Sci 37(02):8–13 (In Chinese). https://doi.org/10.16765/j.cnki.1673-7148.2014.02.016

Yariyan P, Janizadeh S, Van Phong T, Nguyen HD, Costache R, Van Le H, Pham BT, Pradhan B, Tiefenbacher JP (2020) Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping. Water Resour Manag 34(9):3037–3053. https://doi.org/10.1007/s11269-020-02603-7

Zhao G, Pang B, Xu Z, Peng D, Xu L (2019) Assessment of urban flood susceptibility using semi-supervised machine learning model. Sci Total Environ 659:940–949. https://doi.org/10.1016/j.scitotenv.2018.12.217

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.