



UNIVERSIDADE FEDERAL RURAL DO SEMIÁRIDO – UFRSA
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE – UERN
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO – PPgCC
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

RÔMULO GALDINO DA ROCHA LIMA

**DESENVOLVIMENTO DE UM MODELO DE PREVISÃO HIDROLÓGICO
USANDO MÁQUINA VETORIAL DE SUPORTE REGRESSIVO: Uma Aproximação
Computacional para Modelagem da Bacia do rio Arkansas**

MOSSORÓ/RN

2018

RÔMULO GALDINO DA ROCHA LIMA

**DESENVOLVIMENTO DE UM MODELO DE PREVISÃO HIDROLÓGICO
USANDO MÁQUINA VETORIAL DE SUPORTE REGRESSIVO: Uma Aproximação
Computacional para Modelagem da Bacia do rio Arkansas**

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Ciência da Computação – associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semiárido como requisito final para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Dario José Aloise – UERN.

Coorientador: Harold Ivan Ângulo Bustos – UERN.

MOSSORÓ/RN

2018

© Todos os direitos estão reservados a Universidade do Estado do Rio Grande do Norte. O conteúdo desta obra é de inteira responsabilidade do(a) autor(a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu(a) respectivo(a) autor(a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

Catálogo da Publicação na Fonte.
Universidade do Estado do Rio Grande do Norte.

L732d Lima, Rômulo Galdino da Rocha
DESENVOLVIMENTO DE UM MODELO DE
PREVISÃO HIDROLÓGICO USANDO MÁQUINA
VETORIAL DE SUPORTE REGRESSIVO: Uma
Aproximação Computacional para Modelagem da Bacia do
rio Arkansas. / Rômulo Galdino da Rocha Lima. - Mossoró,
2018.
100p.

Orientador(a): Prof. Dr. Dario José Aloise.
Coorientador(a): Prof. Dr. Harold Ivan Angulo Bustos.
Dissertação (Mestrado em Programa de Pós-
Graduação em Ciência da Computação). Universidade do
Estado do Rio Grande do Norte.

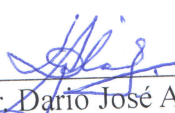
1. Recursos Hídricos, Predição, Redes Neurais, Árvore
de Regressão, SVR. I. Aloise, Dario José. II. Universidade
do Estado do Rio Grande do Norte. III. Título.

RÔMULO GALDINO DA ROCHA LIMA

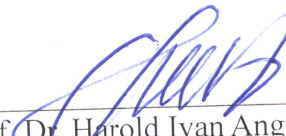
Desenvolvimento de um Modelo de Previsão Hidrológico Usando Máquina Vetorial de Suporte Regressivo: Uma Aproximação Computacional para Modelagem da Bacia do Rio Arkansas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

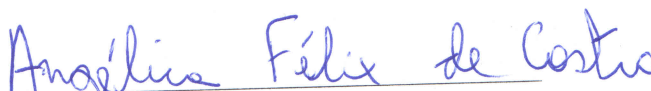
APROVADA EM: 29 / 08 / 2018



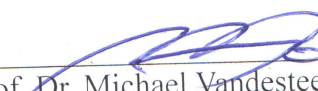
Prof. Dr. Dário José Aloise
Orientador e Presidente



Prof. Dr. Harold Ivan Angulo Bustos
Universidade do Estado do Rio Grande do Norte



Profa. Dra. Angélica Félix de Castro
Universidade Federal Rural do Semi-Árido



Prof. Dr. Michael Vandesteem Silva Souto
Universidade Federal do Rio Grande do Norte - UFRN

AGRADECIMENTOS

Agradeço a Deus por ter me dado a oportunidade de fazer uma pós-graduação, abrindo uma nova oportunidade de pesquisas para enriquecimento do meu conhecimento e das demais pessoas que farão uso deste trabalho.

À minha esposa que, por horas afinco, incentivou-me na melhoria deste trabalho, bem como por todo esforço e ajuda necessários, além da paciência depositada.

Ao professor Dario José, pela orientação, correções e discussões para melhorias desta pesquisa.

Ao professor Harold, pela orientação, paciência e incentivo durante esses anos da pesquisa.

Aos secretários do PPgCC, Rosita e Maninho, por todos os esclarecimentos de dúvidas que surgiram durante o mestrado.

A todos os amigos, companheiros e colegas que surgiram durante esses anos de estudo acadêmico.

*“Felizes aqueles que se divertem com problemas que educam
a alma e elevam o espírito.”*

Fenelon

RESUMO

Os investimentos em meios hídricos naturais são utilizados com finalidades distintas entre os órgãos governamentais que gerenciam esses recursos. Entre tais finalidades, destacam-se o abastecimento de água, a geração de energia, a irrigação e a navegação. A predição da acurácia de tais recursos naturais pode ter grande impacto e relevância para a sociedade, haja vista que a estimativa precisa do comportamento futuro desses recursos pode orientar as autoridades no processo da tomada de decisão, para melhor direcionar os investimentos na prevenção e na mitigação dos efeitos causados pela falta desses recursos, faz-se necessário o desenvolvimento de métodos confiáveis de previsão. Esta dissertação tem por objetivo principal a implementação, aplicação e análise do algoritmo SVR (Support Vector Regression) com a finalidade de se ter um modelo mais preciso computacional de previsibilidade hídrica em relação a outros dois modelos de prognóstico futuro (rede neural e árvore de regressão). Utilizando métricas globais MAE, RMSE e locais, para a exatidão dos modelos, este método proposto aplicou modelos predictivos de imagens, visa ponderar mudanças futuras dos recursos hídricos usando informações obtidas a partir do conjunto dos dados históricos NDVI. As imagens obtidas mostraram as pequenas modificações no intervalo de tempo avaliado de 16 dias para o rio Arkansas. Comparando a precisão dos três métodos, e seus resultados mostraram que o algoritmo SVR conseguiu quantificar de forma mais precisa, estatisticamente, na comparação com outros dois modelos, e ser mais robusto em desempenho computacional para predizer movimentos de fluídos na bacia do rio Arkansas.

Palavras-chave: Recursos Hídricos, Predição, Redes Neurais, Árvore de Regressão, SVR.

ABSTRACT

Investments in natural water resources are used for different purposes among the governmental bodies that manage these resources. Among these purposes are water supply, power generation, irrigation and navigation. The prediction of the accuracy of such natural resources can have a great impact and relevance for society, since the precise estimation of the future behavior of these resources can guide the authorities in the decision-making process, in order to better direct investments in prevention and mitigation of the effects caused by the lack of these resources, it is necessary to develop reliable forecasting methods. This dissertation has as main objective the implementation, application and analysis of the algorithm SVR (Support Vector Regression) in order to have a more precise computational model of water predictability in relation to other two models of future prognosis (neural network and regression tree). Using global metrics MAE, RMSE and local, for the accuracy of the models, this proposed method applied predictive models of images, aiming to consider future changes of water resources using information obtained from the set of NDVI historical data. The images obtained showed the small modifications in the evaluated time interval of 16 days for the Arkansas river. Comparing the accuracy of the three methods and their results showed that the SVR algorithm was able to quantify more accurately statistically in comparison with two other models and to be more robust in computational performance to predict fluid movements in the Arkansas river basin.

Keywords: Water Resources, Prediction, Neural Networks, Regression Tree, SVR.

LISTA DE FIGURAS

Figura 01 - Componentes de um Sistema Sensor	09
Figura 02- Imagem do satélite AQUA com realce no sensor MODIS.....	11
Figura 03 - Etapas do Processamento Digital de Imagens	12
Figura 04 - Elementos da representação vetorial	15
Figura 05 - Processo de Binarização. a) Imagem em escala de cinza. b) Imagem binarizada..	17
Figura 06 - Função de aproximação ϵ – SVR (lado esquerdo), gráfico função de perda ϵ – insensível (lado direito)	22
Figura 07 - Exemplo de uma Árvore de Regressão	25
Figura 08 - Exemplo de um caminho percorrido	26
Figura 09 - Representação de uma Rede de Neurônios Artificiais	28
Figura 10 - Modelo de um neurônio	29
Figura 11 - Região da cidade de Fort Smith e rio Arkansas	31
Figura 12 - Rio Arkansas em local recortado para estudo e visto pelo Fort Smith	32
Figura 13 - Série temporal de imagens da área do Texas e regiões envolventes	34
Figura 14 - Região do rio Arkansas escolhido para estudo	34
Figura 15 - Série Temporal de imagens recortadas do Rio Arkansas River	35
Figura 16 - Série temporal de imagens sintéticas com as seguintes figuras geométricas	38
Figura 17- Interface do Gimp	39
Figura 18- Antes e depois da binarização com realce no rio Arkansas	40
Figura 19- Interface do Qgis 2.18.18	40
Figura 20 - Complementos – Gerenciador e instalação de complementos do Plugin (SCP)	41
Figura 21 - Obtenção de Imagem	41
Figura 22 - Interface do Qgis no processo de classificação	42
Figura 23 - Incersão e Processo de Classificação	42
Figura 24 - Processo de seleção de amostras aleatórias com fins de estimativa da acurácia classificação	43
Figura 25 - Comandos da identificação dos valores dos pixels	44
Figura 26 - Coleta da amostra classificada	44
Figura 27 - Tabela de valores das áreas classificadas	44

Figura 28 - Imagens resultantes da predição de pixel feita pela rede neural e pela árvore de regressão e SVR da 11ª sequência. Simulações do domínio de espaço	47
Figura 29 - Imagens resultantes da predição de pixel feita pela rede neural e pela árvore de regressão e SVR da 12ª sequência. Simulações do domínio de espaço	47
Figura 30 - Imagens resultantes da predição de pixel feita pela rede neural e pela árvore de regressão e SVR da 13ª sequência. Simulações do domínio de espaço	48
Figura 31 - Imagens resultantes da predição de pixel feita pela rede neural e pela árvore de regressão e SVR da 14ª sequência. Simulações do domínio de espaço	48
Figura 32 - Imagens resultantes da predição de pixel feita pela rede neural e pela árvore de regressão e SVR da 15ª sequência. Simulações do domínio de espaço	48
Figura 33 - Imagens resultantes da predição de pixel feita pela rede neural da 13ª sequência. As simulações do domínio de espaço. A fusão da imagem real com as imagens preditas destaca as diferenças nas cores verde e magenta	51
Figura 34 - Imagens resultantes da predição de pixel feita pela rede neural da 14ª sequência. As simulações do domínio de espaço. A fusão da imagem real com as imagens preditas destaca as diferenças nas cores verde e magenta	51
Figura 35 - Imagens resultantes da predição de pixel feita pela rede neural da 15ª sequência. As simulações do domínio de espaço. A fusão da imagem real com as imagens preditas destaca as diferenças nas cores verde e magenta	52
Figura 36 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 11ª imagem	53
Figura 37- Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 12ª imagem	54
Figura 38 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 13ª imagem	55
Figura 39 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 14ª imagem	56
Figura 40 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 15ª imagem	57

LISTA DE TABELAS

Tabela I - Funções Kernel	23
Tabela II - Erro MAE (%) das predições (pixel vs pixel)	50
Tabela III - Erro RMSE (%) das predições (pixel vs pixel)	50
Tabela IV- Erro MAE (%) da predição de pixel das imagens 13, 14 e 15 com a Rede Neural, Árvore de Regressão e o SVR	52
Tabela V- Erro RMSE (%) da predição de pixel das imagens 13, 14 e 15 com a Rede Neural, Árvore de Regressão e o SVR	52
Tabela VI - Porcentagem de Pixels nos mapas temáticos associados aos algoritmos e a 11ª imagem	53
Tabela VII - Porcentagem de Pixels nos mapas temáticos associados aos algoritmos e a 12ª imagem	54
Tabela VIII - Porcentagem de Pixels nos mapas temáticos associados aos algoritmos e a 13ª imagem	55
Tabela IX - Porcentagem de Pixels nos mapas temáticos associados aos algoritmos e a 14ª imagem	56
Tabela X - Porcentagem de Pixels nos mapas temáticos associados aos algoritmos e a 15ª imagem	57

LISTA DE SIGLAS

AM	<i>Aprendizado de Máquina</i>
BRDF	<i>Bidirectional reflectance distribution function</i>
CART	<i>Classification and Regression Trees</i>
MAE	<i>Medium Absolute Error</i>
MODIS	<i>Moderate Resolution Imaging Spectroradiometer</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
RBF	<i>Função de Base Radial</i>
RMSE	<i>Root-Mean-Square-Error</i>
RNA	<i>Rede Neural Artificial</i>
SCP	<i>Semi Automatic Classification Plugin</i>
SIG	<i>Sistema de Informações Geográficas</i>
SVMs	<i>Support Vector Machine</i>
SVR	<i>Regressão de Vetores de Suporte</i>

SUMÁRIO

1 INTRODUÇÃO	05
2 REFERENCIAL TEÓRICO	08
2.1 GEOPROCESSAMENTO.....	08
2.2 SENSORIAMENTO REMOTO - PRINCÍPIOS	09
2.3 Sistemas de estrutura de sensores	09
2.3.1 Sensor MODIS.....	10
2.4 PROCESSAMENTO DE IMAGENS DIGITAIS.....	11
2.4.1 Imagens em NDVI	13
2.5 TIPOS DE DADOS DE SENSORIAMENTO REMOTO	15
2.5.1 Representação Vetorial	15
2.5.2 Representação Matricial – Imagens	16
2.5.3 Binarização de imagens em escala de cinza	16
2.6 SÉRIE TEMPORAL E SUA CONTRIBUIÇÃO.....	17
2.6.1 Algoritmos de Regressão de Imagens de Sensoriamento Remoto.....	19
2.6.2 Implementação.....	19
3 APRENDIZADO DE MÁQUINA	20
3.1 MÁQUINAS DE VETORES POR SUPORTE (SVMs)	20
3.2 REGRESSÃO POR VETORES DE SUPORTE (SVR)	21
3.2.1 Função Kernel e o Kernel RBF	23
3.3 ÁRVORE DE REGRESSÃO	24
3.4 REDE NEURAL ARTIFICIAL	27
3.5 PREVISÃO E MUDANÇAS.....	30
4 ÁREAS ESCOLHIDA PARA O ESTUDO	31
4.1 Cidade de Fort Smith	31
4.2 Rio Arkansas	32
4.3 Desenvolvimento e importância dos rios.....	33
4.4 SEQUÊNCIA DE IMAGENS DE SÁTELITE.....	33
5 ANÁLISE DAS IMAGENS POR MEIO DE MÉTRICAS GLOBAIS REGRESSIVAS	36
5.1 METODOLOGIA GLOBAL REGRESSIVA PARA VALIDAÇÃO	36
5.2 IMAGENS SINTÉTICAS	37

5.3 METODOLOGIA LOCAL PARA VALIDAÇÃO	38
5.4 MODELAGEM HIDRICA NAS SÉRIES TEMPORAIS E SEU COMPORTAMENTO...	45
6 ANÁLISES E RESULTADOS	47
6.1 RESULTADOS DAS IMAGENS DE SENSORIAMENTO REMOTO	47
6.2 RESULTADOS DAS IMAGENS SINTÉTICAS	50
6.3 RESULTADOS LOCAIS DAS IMAGENS DE SENSORIAMENTO REMOTO	53
7 CONSIDERAÇÕES FINAIS	59
7.1 APORTE CIENTÍFICO DA DISSERTAÇÃO.....	60
7.2 TRABALHOS FUTUROS.....	61
REFERÊNCIAS	62
APÊNDICES	67
APÊNDICE A – Códigos de Implementação (Rede Neural, Arvore de Regressão, SVR)	67
APÊNDICE B – Artigo Submetido.....	74

1 INTRODUÇÃO

Com o advento da industrialização, os processos produtivos necessitaram, cada vez mais, de fontes de energias poluentes (e.g. petróleo, carvão mineral, biocombustíveis). Com isso o uso de métodos inteligentes torna-se mais preciso no que se diz respeito a preservação de meios renováveis. Nisto, têm-se revelado mudanças notadamente percebíveis, com uma base científica, na detecção e atribuição de causas, entendimento de variabilidade em meios naturais induzida pelos ciclos hidrológicos ou biogeoquímicos globais, modelando estudos de impactos e adaptação para sistemas e setores relevantes como a agricultura e silvicultura, recursos hídricos, biodiversidade e ecossistema, zonas costeiras, cidades, economia e energia renováveis.

O desenvolvimento de conhecimento e as tecnologias para a mitigação das mudanças climáticas têm se mostrado promissores e ganhado destaque em discussões frequentes. Ante a crescente ordem de impactos que atingem o mundo em relação aos seus eventos catastróficos, abre-se a necessidade de entender o porquê tais eventos são mais progressivos em alguns casos (ação humana) que outros (própria natureza).

Este trabalho cria novas perspectivas sobre predição de séries temporais a partir da utilização de imagens de Sensoriamento Remoto por meio da aplicação de técnicas de PDI e de Geoprocessamento. Com este olhar, abre-se uma nova problemática, posto que muitas vezes há a necessidade de tomadas de decisão que possam impactar previamente em algo, seja na prevenção, seja na mitigação da falta de recursos. A predição é feita não somente de um único modelo, mas para que este seja eficaz em todos os tipos de séries.

A princípio, esta pesquisa inicia uma nova técnica de modelagem preditiva de séries temporais para estimar e quantificar os recursos “naturais” a partir de sequências de imagens de sensoriamento remoto. Pretende-se com isso avaliar quantitativamente a estratégia da solução proposta para a resolução do problema da estimação preditiva, computando o erro mínimo quadrado entre o modelo real e o modelo estimado, de forma a definir qual dos dois métodos possuem o menor erro quando aplicada ao problema em questão.

A Regressão por Vetores de Suporte (SVR, do inglês Support Vector Regression) constitui numa técnica de aprendizagem que vem recebendo crescente atenção da comunidade de Aprendizado de Máquina (AM). Os resultados da aplicação dessa técnica são comparáveis e, muitas vezes, superiores aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs) e a Árvore de Regressão. Os exemplos de aplicação de grande sucesso podem ser encontrados em diversos domínios, como a análise de imagens.

Modelos de previsão possuem várias opções na sua fundamentação e definição de comportamento. Por meio deles, determinam-se parâmetros que servem para avaliar seu melhor desempenho. Desta forma, é preciso que o modelo utilizado apresente um alto grau de confiabilidade para a obtenção de resultados confiáveis e satisfatórios.

Tendo em vista os bons resultados apresentados pelo algoritmo Regressão de Vetores de Suporte (SVR - *Support Vector Regression*) por pertencer a teoria de aprendizado estatístico e trabalhar com ferramentas de preditivas, mostrou sua eficácia em diversos campos do conhecimento como medicina, inovações tecnológicas no que se diz respeito à previsão de imagens (YIGIT *et al.*, 2018; FENTIS *et al.*, 2017) esta técnica torna-se muito atrativa e uma ótima alternativa para obtenção de um modelo de previsão em regressão linear de qualidade. Ademais, importante ressaltar que tal tema é encontrado na literatura com outros casos de estudo, no que concerne à utilização do algoritmo SVR. A presente dissertação trata desta matéria, além de alcançar resultados com outros algoritmos utilizados (rede neural e árvore de regressão).

Esta pesquisa tem como objetivo principal adaptar, aplicar e avaliar a técnica preditiva do algoritmo de Regressão de Vetores de Suporte (SVR) em séries temporais de imagens NDVI do sensor MODIS, para o monitoramento processadas digitalmente com uso da técnica de Normalized Difference Vegetation Index (NDVI). De forma mais analítica, avalia quantitativamente a estratégia da solução proposta pelo SVR para a resolução do problema da estimação preditiva entre o modelo real e o estimado, comparando a eficácia do método SVR a partir da utilização de dados de séries temporais das imagens NDVI com outras técnicas de predição, aplicado na modelagem preditiva de recursos hídricos no solo.

Este trabalho está dividido em seis capítulos, além da parte introdutória e das considerações finais.

No segundo capítulo, faz-se um aparato acerca das noções de sensoriamento remoto e sua estruturação de interpretação. Ainda, faz a citação de trabalhos relacionamentos a sensoriamento remoto com imagens em NDVI e a importância do uso do geoprocessamento no nosso caso de estudo.

O terceiro capítulo traz conceito do aprendizado de máquinas e os algoritmos de inteligência computacional que serão utilizados, assim como a importância de utiliza-los.

No quarto capítulo mostra a área escolhida para o estudo de caso, visto que o foco são as mudanças hídricas envolventes do rio Arkansas.

O capítulo cinco, traremos algumas métricas de regressão para analisar estas imagens de sensoriamento remoto, assim com estratégias de predição de séries temporais com métricas locais e globais e utilização de imagens sintéticas para validação.

No sexto capítulo os resultados utilizados para as estratégias de análise das imagens de sensoriamento remoto em NDVI.

Almeja-se que o referido trabalho proporcione ao leitor um alcance de informações de modo que possam esclarecer um pouco sobre esta temática, além de que permita entender como mudanças bruscas no ambiente pode afetar seu dia-a-dia e gerar acontecimentos importunos a sobrevivência humana da fauna e flora.

2 REFERENCIAL TEÓRICO

As várias modificações que ocorrem no nosso planeta muitas vezes é fruto de ocorrências naturais e pela ação humana, nisso foram criados mecanismos e ferramentas que possam ajudar ao homem a entender seus efeitos benéficos ou de grande mal ao transcorrer dos dias e anos. Serão vistos um pouco desses mecanismos de monitoramento que circundam os estudos científicos e sua previsibilidade de acontecimentos.

2.1 GEOPROCESSAMENTO

O estudo sobre os pontos de classificação geográfica que envolvem nosso planeta sempre foi algo a ser explorado por pesquisadores da área, uma vez que documentos que envolvam os recursos existentes, assim como fauna, flora e estruturação social, motivaram estes estudiosos a aprofundarem-se e a buscar ferramentas que auxiliem as suas pesquisas. Com a evolução tecnológica, novos caminhos e possibilidades de ajuda abriram-se nas pesquisas, utilizando-se de ambientes virtuais com tecnologias inteligentes que pudessem reproduzir tais ideias, surgindo assim o termo Geoprocessamento, que nada mais é que um tratamento de dados geográficos gerando diversas informações no meio topológico e georreferenciamento, utilizado para fins de pesquisa e análise (CÂMARA; DAVIS; MONTEIRO, 2001).

Muitos estudiosos do ramo utilizam o termo de Geoprocessamento para descrever o uso de metodologias ou procedimentos que envolvam a matemática e a computação em conjunto para adquirir informações geográficas para diversos fins, onde envolvam melhorias e compreensão de acontecimentos local, regional, global tudo em questão de escala, sejam eles diários, mensais, anuais ou até mesmo decadais. A utilização do geoprocessamento como técnica e metodologia, fornece ferramentas para a análise de suas complexidades, que são Sistemas de Informação Geográfica (SIG), os SIG produzem vários bancos de dados que envolvem o georreferenciamento de vários tipos de imagens e fontes para a confecção de documentos sobre um determinado estudo a ser analisado (MARQUES FILHO; VIEIRA NETO, 1999; CÂMARA; DAVIS; MONTEIRO, 2001).

2.2 SENSORIAMENTO REMOTO - PRINCIPIOS

Kumar, Ali e Hammed (2012), afirmaram que a radiação eletromagnética, ao precipitar-se sobre a superfície de um determinado material em partes, será refletida por esta superfície, em outras, absorvida, e, em outra parte, será propagada caso a matéria possua alguma translucidez.

As definições para Sensoriamento remoto são inúmeras deste o seu princípio clássico até o científico, saindo de ser uma ferramenta para uma ciência, assim ela é uma ciência que propõem-se na aquisição de imagens da superfície terrestre por meio verificação de medidas e testes da influência de radiação eletromagnética e objetos terrestres (MENEZES; ALMEIDA, 2012).

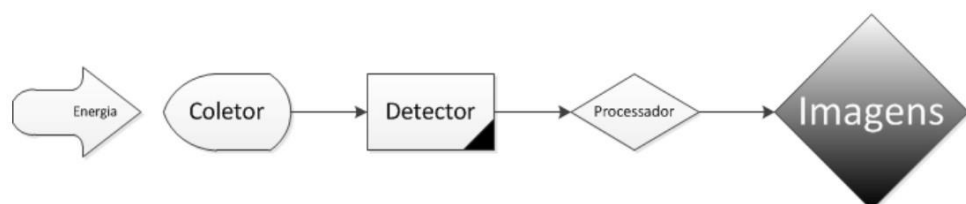
Tendo esse princípio temos a visão que objetos terrestres monitorados por sensores por meio de emissão de radiação eletromagnética, parte do meio que objetos que são refletidos pelo sol propagam esse local na superfície terrestre. Se não tiver uma captação dessas imagens por sensores é desconsiderada o princípio do sensoriamento remoto. Assim os sensores remotos são ferramentas de total importância para a realização de pesquisas envolvendo mapeamento e monitoramento de recursos naturais e do seu meio.

2.3 Sistemas de estrutura de sensores

Os sistemas sensores são dispositivos capazes de detectar e registrar a radiação eletromagnética em determinada faixa do espectro eletromagnético e gerar informações que possam ser transformadas num produto passível de interpretação, na forma de gráficos ou de tabelas (MOREIRA, 2005).

Na Figura 01, o coletor recebe a energia através de uma lente e passa a informação para o detector que capta a energia coletada de uma determinada faixa do espectro. De posse dessas informações, o processador executa o sinal registrado através do qual se sintetiza uma imagem.

Figura 01 – Componentes de um Sistema Sensor.



Fonte: Moraes (2002).

Energia - energia eletromagnética refletida para a direção dos sensores;

Coletor - Componente óptico que tem a função de concentrar a energia que vem da amostra coletada;

Detector – Componente com pequenas dimensões, cujo propósito é absorver fluxo de energia que é gerado por sinais elétricos;

Processador – Componente que tem a propriedade de amplificar sinais fracos que são produzidos pelo detector.

Processos que envolvem gerar dados a partir de imagens adquiridas pela conversão da intensidade luminosa em sinais eletromagnéticos, são existentes em vários sensores, pois foram instalados em satélites para obtenção de imagens aéreas, esses sensores que fazem a cobertura do terreno deve ser feito de forma linear a se fazer a varredura ou cobertura captada pelo sensor em varias faixas de espectro eletromagnético simultâneos (MENEZES; ALMEIDA, 2012).

O sensor que foi utilizado para a obtenção das imagens processadas em NDVI na pesquisa dessa dissertação foi o MODIS do satélite AQUA, onde será explorado seus detalhes no item a seguir.

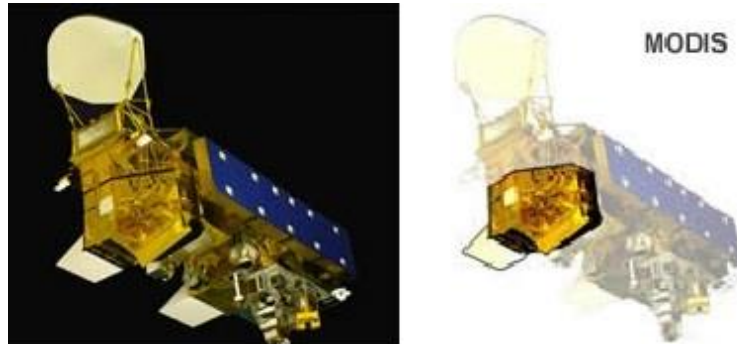
2.3.1 Sensor MODIS

O sensor MODIS é um instrumento desenvolvido pela NASA como componente do satélite AQUA, um satélite americano desenvolvido em parceria com o Japão e o Brasil para fenômenos que envolvem a circulação de energia e da água na superfície terrestre e que faz operação em 36 canais espectrais que vão de 0,4 a 14,4 μm e resolução espacial que vai de 250 a 1.000 metros (LATORRE *et al.*, 2003). As imagens geradas pelo MODIS apresentam uma base de dados ordenada na elaboração de resultados envolvendo atmosfera, superfície terrestre e oceanos. As aplicações são inúmeras para a utilização desse sensor como em exemplos temos:

- Adquire medidas de fluxo de energia radiante;
- Medidas e monitoramento de dados envolvendo mudanças climáticas globais;
- Atividades de uso da cobertura das terras;
- Medidas que envolvem umidade, temperatura da superfície marítima e terrestre.

No satélite AQUA o sensor MODIS é o que se destaca entre os seis sensores a bordo do satélite (Figura 02) por ter sua eficiência singular envolvendo trabalhos que envolvam agricultura devido a sua resolução temporal que permite monitorar sistemas de cultura agrícola como colheitas e recursos que envolvem essas atividades.

Figura 02 – Imagem do satélite AQUA com realce no sensor MODIS.



Fonte: NASA (2007).

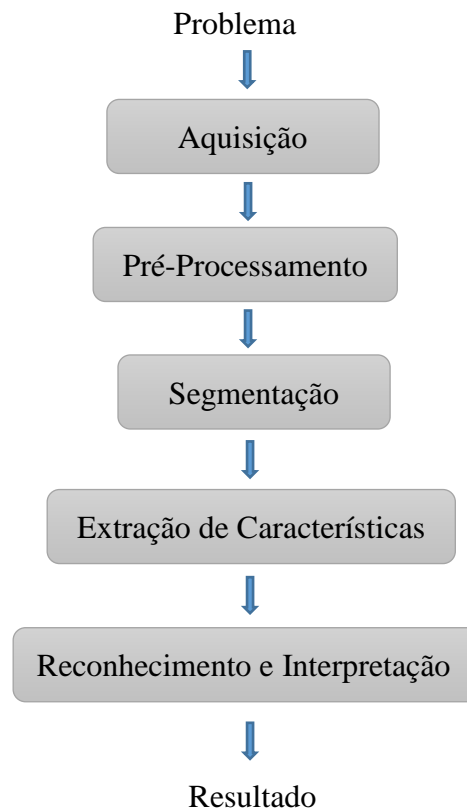
As imagens geradas por este sensor foram utilizadas neste trabalho com processamento em NDVI, visando ponderar mudanças hídricas envolvendo a bacia do rio Arkansas, através da utilização de algoritmos de aprendizado de máquina.

2.4 PROCESSAMENTO DE IMAGENS DIGITAIS

A obtenção de informações a partir de geoprocessamento, muitas vezes, requer técnicas de melhoria dos dados para utilização em determinados fins ou locais que se pretende fazer uso. Nesse sentido, utilizam-se procedimentos para se tratar esses dados.

Processamento Digital de Imagens (PDI), trata-se técnicas e métodos utilizados para realçar as informações, possibilitando uma melhor individualização das unidades mapeadas por meio do tratamento de imagens com a utilização de procedimentos computadorizados pré-definidos ou manipulados por ajustes no histograma da imagem e aplicações de novos algoritmos no entender do próprio usuário (QUEIROZ; GOMES, 2001). O PDI está dividido nas seguintes etapas como nos mostra a Figura 03:

Figura 03 – Etapas do Processamento Digital de Imagens.



Fonte: adaptado, Marques Filho; Vieira Neto (1999).

Queiroz e Gomes (2001) explicam as etapas da seguinte forma:

Etapa de Aquisição: coleta e obtenção de imagens criando modelos da cena real, bem como a obtenção dessa imagem. Essa etapa tem um papel fundamental, que é o de gerar dados numéricos para que depois possam ser computados.

Etapa de Pré-Processamento: em muitos casos, é preciso resolver certos problemas depois que acontece a aquisição, como ruídos (falhas) adquiridos pelos sensores dos satélites. Essas distorções podem ser corrigidas e ponderadas. Para tanto, essa etapa também tem seu valor, posto que abre a interpretação para novos elementos posteriores que venham a aparecer.

Etapa de Segmentação: é a etapa onde ocorre a divisão da imagem. É uma etapa complexa de se formalizar, pois ela vistoria caracteres e sua similaridade entre os elementos.

Extração: pondera os caracteres dos elementos após a segmentação.

Etapa de Reconhecimento e interpretação: não menos importante, é, de fato, reconhecida como a etapa de titular os caracteres. Ainda, nessa etapa dá-se o entendimento que os elementos titulados são utilizados em diversas áreas do conhecimento humano e na análise computadorizada.

Nas etapas citadas acima, quanto ao tratamento de imagens, elas podem ser utilizadas para distintas metodologias e técnicas de modo prático.

No nosso caso de estudo, a utilização das imagens em escalas de cinza monocromáticas promove o trabalho com processos para a separação ou divisão de elementos existentes nas imagens, onde estes elementos podem ser mantidos ou destacados em seus caracteres de tonalidade. Essa técnica de segmentação em tons de cinza é viável e tem grande valor nas pesquisas de PDI, uma vez que trabalhar com essas imagens é mais hábil que trabalhar com imagens que envolvem coloração, devido ao processamento e a utilização de algoritmos simples, além de utilizar menos recurso em termos computacionais (KANAN; COTTRELL, 2012).

2.4.1 Imagens em NVDI

O Índice de Vegetação da Diferença Normalizada (NDVI) é um índice que permite identificar a presença de vegetação verde na superfície e caracterizar sua distribuição espacial. Essas informações combinadas podem ser muito importantes para identificar fenômenos que podem estar ocorrendo em uma determinada área, notadamente os relacionados com os processos de degradação (SILVA *et al.*, 2009).

O NDVI é uma técnica de processamento de imagens que viabiliza a otimização e análise de dados que envolvam o sensoriamento remoto de imagens, fornecido pela razão entre as bandas do vermelho e do infravermelho próximo. Tratando-se de uma habilidade matemática que utiliza a refletância dos alvos nessas faixas do espectro com o fim de se obter um produto no qual demonstre e mostre alvos de vegetação sadia ressaltadas pelas maiores refletância, neste caso exposto em tonalidades mais claras na escala de cinza. Isso só é possível pois a vegetação apresenta refletância baixa ou seja na faixa visível (em função de absorver a radiação eletromagnética pelos pigmentos de fotossíntese ativos na folha) e também a alta no infravermelho que está próximo (devido a reprodução da radiação eletromagnética da construção interna da folha) (MOREIRA, 2005).

Os índices de vegetação foram fundamentados afim de diminuir o número de parâmetros contidos nas medidas multiespectrais. Esses índices são produzidos a partir de dados de sensoriamento remoto, e sua constituição é de extrema importância estratégica para acompanhar as alterações naturais ou produzidas pelo homem no ecossistema e sua biomassa.

Os valores do NDVI oscilam entre -1 e +1, sendo valores negativos correspondentes à presença de água (rios, lagoas, etc.) e valores positivos representam uma vegetação exuberante, com grande quantidade de biomassa. Este índice permite identificar a presença verde na superfície e caracterizar sua distribuição espacial, assim como sua evolução no decorrer do tempo (SILVA *et al.*, 2009), cujo objetivo é eliminar diferenças de determinada época ou ângulo do sol e diminuir os efeitos da redução atmosférica, proporcionando uma normalização com a qual os valores mostrados são obtidos e contidos em uma mesma escala (adimensional) de valores.

A equação do NDVI é expressa pela seguinte fórmula:

$$NDVI = \frac{P_{NIR} - P_{RED}}{P_{NIR} + P_{RED}} \quad (1)$$

Onde contém nessa equação:

NDVI - Índice de vegetação por diferença normalizada;

P_{NIR} - Refletância de superfície no comprimento de onda do infravermelho;

P_{RED} - Refletância de superfície no comprimento de onda vermelho.

As características temporais foram detectadas a partir de 15 imagens do satélite AQUA do sensor MODIS, com 250 m de resolução espacial, Rouse *et al.* (1973). As imagens do sensor MODIS foram disponibilizadas pelo professor Byron Tapley, diretor chefe do Centro de Pesquisas Espaciais da Universidade do Texas, USA, onde cada imagem processada mostra um mosaico com constituição de *pixels* das imagens diárias de refletância de superfície no período de 16 dias sucessivos, compreendidos entre os meses de maio a dezembro de 2002.

A constituição das imagens em NDVI são geradas utilizando-se a função de refletância bidirecional (BRDF) de Walthall. Os parâmetros desta função são utilizados para normalizar e interpolar a refletância de superfície no ângulo de visão do qual são registrados os valores de NDVI da série para um determinado *pixel*. Assim, fatores como nebulosidade, variações do ângulo de iluminação e geometria visada, vapor de água, aerossóis etc., são minimizados.

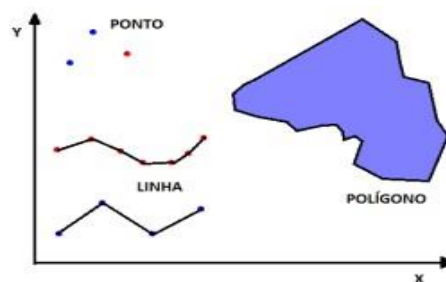
2.5 TIPOS DE DADOS DE SENSORIAMENTO REMOTO

O uso de dados de sensoriamento remoto viabiliza um conjunto de dados para obtenção de informações que compõem a superfície terrestre, assim estudar sua estruturação em formação e interpretação, pondera resultados satisfatórios refletidas por objetos terrestres registrados.

2.5.1 Representação Vetorial

Câmara, Davis e Monteiro (2001) definiram o conceito de representação vetorial como a forma reduzida de um elemento ou objeto. Esse elemento pode ser representado de três formas básicas: pontos, linha poligonal, e áreas ou polígonos, como pode-se visualizar na Figura 4:

Figura 04 – Elementos da representação vetorial.



Fonte: Câmara; Davis; Monteiro (2001).

Onde:

- Ponto – será um par ordenado (x, y) de coordenadas espaciais. Outros dados não espaciais, denominados atributos, também podem ser arquivados para indicar características do ponto;
- Linha Poligonal – consiste em um conjunto de pontos conectados. Nas linhas, são armazenadas, além das coordenadas dos pontos, informações com a indicação do tipo de linha que está sendo tratada, ou seja, a qual atributo ela está associada;
- Áreas ou Polígonos – correspondem à região do plano, limitada por uma ou mais linhas poligonais conectadas de modo que o último ponto de uma linha seja o primeiro ponto da próxima linha.

2.5.2 Representação Matricial – Imagens

As imagens digitais da superfície terrestre gravadas por sensores orbitais, como o sensor TM do LANDSAT e o HRV do Spot, são transformadas em dois tipos de produtos: analógico e digital (MOREIRA, 2005).

Os produtos analógicos são tratados pelo processo de interpretação visual. Os produtos digitais são tratados por meio de métodos específicos de análise de dados e implementados em computadores, passando essa imagem por um processo de melhoramento de alguns aspectos de suas feições.

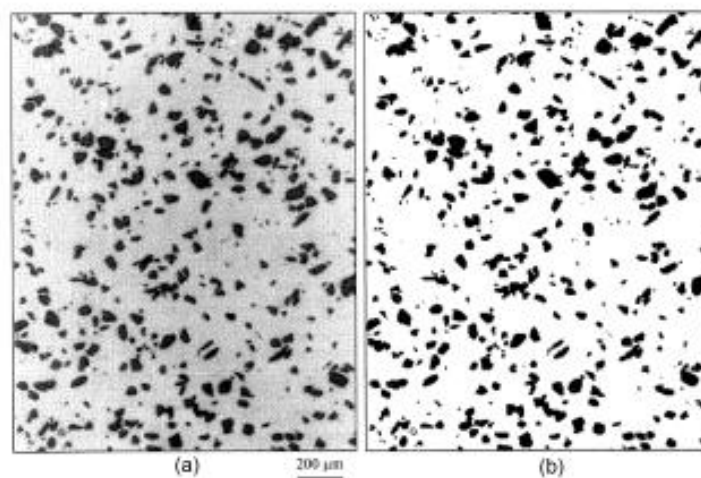
O satélite AQUA com o sensor MODIS (Moderate Resolution Imaging Spectroradiometer), segundo Barker *et al.* (1992) apresenta variadas características, entre se destacam:

- Cobertura extensa tanto espacial como espectral;
- Adquire medições em locais espectrais que disponibiliza sistemas de sensores com conjunto de dados, sendo utilizado para meteorologia e obtenção de temperatura aquática.
- Sendo utilizada como ferramenta para direção de pesquisas em mudanças e não mudanças no planeta.

2.5.3 Binarização de Imagens em Escala de Cinza

Binarização de imagens é uma metodologia utilizada na segmentação de elementos existentes nas imagens em escala de cinza em distintas condições ou níveis que esta imagem se encontre. Trabalhar com essa metodologia resulta em adquirir dados quantitativos de importância para o objeto em estudo (PAL; PAL, 1993). A binarização aplica a propagação ou distribuição da amplitude de *pixels* que estejam bem diferentes entre si, fazendo a divisão em duas partes, onde uma é a parte que se tem interesse em investigar e a outra será o fundo, partindo de um princípio no qual, segundo alguns autores, segmentam em duas regiões de cor branca e preta, em que uma é o foco de estudo e a outra é o fundo, ou vice-versa (PAL, PAL, 1993; SENTHILKUMARAN; VAITHEGI, 2016), como nos mostra a Figura 5:

Figura 05 - Processo de Binarização; a) Imagem em escala de cinza; b) Imagem binarizada.



Fonte: Silva (1996).

O processo que resulta numa imagem binária é chamado de binarização, delimitando os valores dos pixels em camadas ou grupos. Essa técnica é muito conhecida e utilizada para o estudo de feições e características e, devido ao seu proveito, mostra elementos e uma visão computacional quantitativa de objetos que possam aparecer de modos semelhantes ou diferentes nos modelos em estudo (RAJAGOPALAN *et al.*, 2004).

A transição das imagens em escala de cinza que passam para um processo de binarização é uma técnica presente utilizada em áreas de estudo como geologia, medicina e estudo de certos materiais para confecção de objetos, posto que os realces mostram soluções visíveis para análise e estudo. Portanto, essa forma de distinguir as tonalidades dos elementos envolvidos na imagem faz com que a obtenção de resultados que utilizam a binarização nas imagens em escala de cinza seja mais proveitosa que a coleta de bons resultados (RAJAGOPALAN *et al.*, 2004).

2.6 SÉRIE TEMPORAL E SUA CONTRIBUIÇÃO

Segundo Cowpertwait e Metcalfe (2009), a fundamentação de série temporal é dada como coleta de observações de uma determinada variável, sendo que essas observações são periódicas, durante um certo intervalo de tempo, onde alguns pesquisadores denominam como período amostral.

A série temporal tem como objetivo as observações feitas durante um certo intervalo de tempo e é representado pela seguinte expressão: $X_l = \{ x_1, x_2, \dots, x_n \}$ equivale a x , sendo este a variável dos testes e n a quantidade de testes ou medições adquiridas (SHUMWAY;

STOFFER, 2011). Para cada série temporal, temos vários campos e casos de estudo onde ela pode ser usada como exemplificação, por exemplo:

- Temperatura local e global;
- Indicadores inflacionários de economia;
- Alíquotas de emprego e desemprego;
- Indicadores hidrológicos e pluviométricos.

Muitos dos exemplos citados são feitos através de estudos que utilizam séries temporais para a elaboração, análise e estudo referente às mudanças ocorridas ao longo do tempo, pois considera-se que valores quantitativos previstos podem ter origem de sua distribuição, a qual se saiba sua média e as variações ocorridas (COWPERTWAIT; METCALFE, 2009).

Nas séries temporais existem uma convergência ou tendência de caracteres, que se repetem com o passar do tempo, existindo assim uma sequência de valores na série em estudo e a ocorrência de diminuição ou aumento da média desses valores periódicos da série (COWPERTWAIT; METCALFE, 2009).

Alguns preditores especializados na padronização desses comportamentos de série temporal procuram sempre uma melhoria na precisão ou exatidão de resultados. A previsão utilizando séries temporais tem como foco a medição de valores futuros, fornecidos antes pelo seu passado, provendo informações externas que não estão contidas na série (BOSER, GUYON; VAPNIK, 1992; CORTES, VAPNIK, 1995). Esse tipo de estudo tem sua importância na contribuição como auxiliar na tomada de decisões, como o exemplo citado das indicações pluviométricas: faz-se uma estimativa, caso o valor seja baixo, tem-se uma indicação que futuramente será preciso um controle de distribuição de águas, comumente conhecido como racionamento, pela população ao qual é fornecido esse serviço; no entanto, se os valores forem altos, podem trazer danos em decorrência do excesso de água, provocando alagamentos, impactando em ruínas de casas, prédios e até plantações, além de acarretar certas doenças à população. Ante o exposto, estudar as séries temporais moldadas com preditores especializados podem trazer vantagens para a tomada de decisão a serem adotadas pelo governo da região objeto do estudo (COWPERTWAIT; METCALFE, 2009).

2.6.1 ALGORITMOS DE REGRESSÃO DE IMAGENS DE SENSORIAMENTO REMOTO

Desde o advento da Inteligência Artificial, os algoritmos de aprendizagem de máquinas vêm sendo utilizados, ganhando mais força depois dos anos 80, para fazer operações referentes a Classificação e a Predição (FACELI *et al.*, 2011; BARBOSA, 2013).

O foco deste trabalho recairá sobre a operação de predição, para fazer extrapolações de séries temporais, visando melhorias tanto na análise como no tratamento dessas imagens. Correlacionado com o aprendizado de máquina, os conceitos e algoritmos referentes a esse tema descreve como o aprendizado preditivo envolvendo (Redes Neurais, Árvore de Regressão e o SVR) combinados podem formar comitês para planejar, experimentar e empregar métricas para avaliação de desempenho

2.6.2 Implementação

Para implementar os algoritmos (SVR, Árvore de Regressão e Rede Neural) foi utilizado o software Matlab que nada mais é que uma linguagem apropriada de programação no intuito de desenvolver de maneira técnica aplicações nas mais variadas pesquisas que envolvam precisão sem perda de tempo com detalhes. De tal modo, ajuda na ampliação dos campos de programação e visualização de soluções de forma amigável ao aprendizado (GILAT, 2012).

A origem do seu nome vem de MAT (*Matrix*) e LAB (*Laboratório*) um variado laboratório virtual de programação que consiste no uso de matrizes e vetores para o estudo daqueles que desejam implementar e fazer os testes de suas soluções (GILAT, 2012).

No prezado modo, esse ambiente de desenvolvimento dispõe de bibliotecas que auxiliam na geração de resultados envolvendo funções matemáticas, desenhos e gráficos manipulados que ajudam na interpretação dos trabalhos. Contendo ainda com auxílio para determinadas áreas de estudo um campo de bibliotecas contendo EDO (*Equações Diferenciais Ordinarias*) e campos visuais como processamento de imagens e sinais e o campo estatístico (GILAT, 2012).

Assim com esta vasta variedade, os usuários que fizerem uso de tal ferramenta terão um grande apoio lógico no que se diz respeito as pesquisas no campo de vetores e matrizes (GILAT, 2012).

3 APRENDIZADO DE MÁQUINA

Com o avanço das décadas, novas metodologias do uso da inteligência artificial foram formuladas com o intuito de ajudar o ser humano na resolução de problemáticas do seu cotidiano como (1) sistemas individualizado e (2) sistemas com apoio de históricos passados de amostras ou modelos do conhecimento humano (FACELI *et al.*, 2011; BARBOSA, 2013).

Instrumentos de aprimoramento fazem com que essa metodologia possa resolver problemas fáceis e difíceis na tomada de solução, mas, mesmo assim, observa-se a produção de informações em vários campos de atuação humana dificultando as vezes na exatidão dessa metodologia de inteligência (FACELI *et al.*, 2011; BARBOSA, 2013).

De tal modo, a base do aprendizado de máquina tem como foco a inferência de histórico passado de informações conhecidas, ou seja, sua “indução” (FACELI *et al.*, 2011; BARBOSA, 2013).

3.1 MÁQUINAS DE VETORES POR SUPORTE (SVMs)

Há algum tempo, o número de pesquisas e pesquisadores que fazem o uso do conjunto de métodos do Suporte por Máquinas de Vetores por Suporte (SVMs) obteve um crescimento significativo e com isso a comunidade acadêmica de aprendizagem de máquina passou a dar mais importância. Esses métodos têm, em alguns casos, obtido resultados que muitas vezes são superiores a outros algoritmos famosos de aprendizado de máquina, como a rede neural artificial (SANTOS, 2002; FACELI *et al.*, 2011).

As máquinas de vetores por suporte têm a base fundamentada na teoria pelo aprendizado estatístico, criada por Vapnik em 1995. O SVM apresenta duas características que sobressaem às demais, que são: (1) uma fundamentação teórica consistente; (2) atinge um grande desempenho em aplicações práticas do conhecimento humano. A validação dessa aprendizagem é capaz de reconhecer precisamente as características na qual precisam ser considerados para a aprendizagem ser bem sucedida e o desenvolvimento de modelos que são muito complexos (SANTOS, 2002; BISHOP, 2006; HASTIE *et al.*, 2013).

Na etapa de destreza do SVM é aplicada uma função quadrática convexa, no qual tem como critério levantar uma problemática no que se refere à otimização matemática. Os parâmetros são ajustados pelo usuário e também não há nenhuma submissão entre as partes, esses ajustes são alterados dependendo do contexto em que está inserido o SVM (KWOK, 1998; SANTOS, 2002).

A otimização e aprimoramento matemáticos, que incluem características das soluções para categorias de problemáticas formadas por funções que devem ser escolhidas para minimizar ou maximizar uma certa função custo e algumas restrições, podem ser incorporadas. Otimização, em termos matemáticos, preocupa-se com a performance dos algoritmos na procura por tais respostas ou soluções. Com isso a hipótese que se forma na otimização oferece as técnicas algorítmicas e inclusive define encargos necessários e eficazes para um função dada ser uma possível solução ou resposta a uma certa problemática (SANTOS, 2002; FACELI *et al.*, 2011).

O SVM lida com problemas que envolvem tanto a classificação ou a regressão. No nosso caso de estudo, utilizou-se o SVR (Regressão por Vetores de Suporte), introduzido por Vapnik em 1995, a partir de algumas ideias e pensamentos do algoritmo SVR, no que se refere a buscar uma função $f(x)$ que produza saídas sucessivas e contínuas para um certo dado de treinamento que distancie ao máximo de seus rótulos já estabelecidos (FACELI *et al.*, 2011).

3.2 REGRESSÃO POR VETORES DE SUPORTE (SVR)

O funcionamento do algoritmo SVR *Regressão por Vetores de Suporte* baseia-se na metodologia do aprendizado supervisionado. Tal metodologia faz com que um sistema aprenda uma função a partir de uma amostra de dados de treinamento $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \subseteq (X \times Y)^n$ onde X denomina o espaço de entrada e Y o espaço de saída. Nesse contexto, o objetivo do SVR é encontrar uma função $f(x)$ que apresente, no máximo, um desvio ε em relação aos valores alvo y_i . Em outras palavras, procura-se uma função com uma margem de erros caracterizada pelo intervalo $[y_i - \varepsilon, y_i + \varepsilon]$ e que seja a mais paralela do domínio possível de soluções (SMOLA; SCHÖLKOPF, 2004).

Assumindo funções nessa forma, tem-se:

$$f(x) = \langle w, x \rangle + b \quad (2)$$

Onde: $w \in \mathbb{R}^n, b \in \mathbb{R}$.

Rodrigues *et al.*, (2015) afirma que buscar uma função mais paralela ao domínio possível significa determinar valores pequenos para o vetor dos pesos w , o que corresponde a minimizar a medida $\|w\|^2$ (SMOLA; SCHÖLKOPF, 2004). Desta forma, modela-se o problema de otimização primal do SVR como sendo:

$$\text{Minimizar:} \quad \frac{1}{2} \|w\|^2 \quad (3)$$

$$\begin{aligned} \text{Restrições Sujeitas:} \quad & \{ y_i - w \cdot x_i - b \leq \varepsilon_i \\ & \{ w \cdot x_i + b - \leq y_i \varepsilon_i \end{aligned}$$

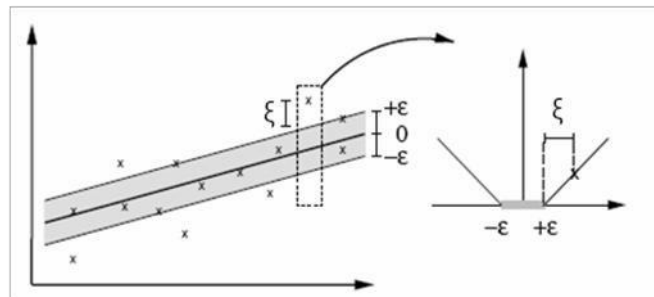
Onde: $w \in \mathbb{R}^n$ e $b \in \mathbb{R}$ são as incógnitas do problema.

Entretanto, nem sempre é possível garantir a viabilidade do problema, tendo em vista que existem pontos que violam suas restrições. Por esta razão, estuda-se uma função de perda, que introduz variáveis de folgas não negativas ξ_i, ξ_i^* , cuja finalidade é penalizar dados que se situem fora da margem $|f(x_i) - y_i| \leq \varepsilon$, (VAPNIK, 1995). Essa função de perda, denominada ε -Insenstitiva, é descrita por:

$$\begin{aligned} |\xi|_\varepsilon &= 0, & \text{se } |\xi| \leq \varepsilon \\ |\xi|_\varepsilon &= |\xi| - \varepsilon & \text{caso contrário} \end{aligned} \quad (4)$$

A Figura 06 representa graficamente a situação proposta pela expressão (4) na qual apenas os pontos localizados fora da região sombreada contribuem para o valor custo da função.

Figura 06 – Função de aproximação ε – SVR (lado esquerdo), gráfico função de perda ε – Insensitiva (lado direito).



Fonte: adaptado, Smola; Schölkopf (2004).

A solução de tais problemas nem sempre é fácil de ser resolvida, em comparação às restrições com sua desigualdade. Assim, esse protótipo passará a ser mostrado num determinado espaço dual, o que proporcionará mais eficácia e agilidade ao algoritmo SVR (FACELI *et al.*, 2011). Descrevendo graficamente a situação, apenas os pontos fora da região sombreada contribuem para o custo, na medida em que os desvios são penalizados de forma linear. Acontece que, na maioria dos casos, o problema de otimização pode ser resolvido mais facilmente diante da formulação dada.

3.2.1 Função Kernel e o Kernel RBF

Conceituando o esquema de uso da teoria de Cover, afirma-se que dado um conjunto de dados linear, ele apresenta maior possibilidade de ser linearmente separável, em um espaço com maior dimensionalidade. Dados as transformações de dimensões e sendo feita através das funções Kernel, conseguindo assim uma problema linear, podendo então aproveitar e encontrar um hiperplano ótimo (FACELI *et al.*, 2011; GONÇALVES, 2015).

A fundamentação teórica nos fala que a função Kernel têm como entrada dois tipos de dados x_i e x_j , onde esses dados geram um produto de modo escalar, onde as características de espaço são predominantes, citado por Faceli *et al.* (2011); Gonçalves (2015), a equação a seguir mostra estas entradas e seu produto escalar.

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (5)$$

Os ganhos de informação são mostrados com a função Kernel e seu mapeamento Φ onde é obtido resultados que envolvem o cálculo de espaço com dados característicos, com esse cálculo é otimizado e colocado certas restrições que possa gerar baixa complexidade nos resultados, porém proporcionalmente também um baixo custo computacional (FACELI *et al.*, 2011; GONÇALVES, 2015).

A tabela I, a seguir mostra os Kernels, assim como seus parâmetros, onde esses parâmetros são identificados como matrizes com valores que ressaltam do custo positivo e construção ainda não definida desses valores (FACELI *et al.*, 2011).

Tabela I: Funções Kernel.

Tipo de Kernel	Função $k(x_i, x_j)$	Parâmetros
Polinomial	$(\delta(x_i, x_j) + k)^d$	δ, K e d
RBF	$\text{Exp}(-y\ x_i - x_j\ ^2)$	y
Sigmoidal	$\tanh(\delta(x_i, x_j) + k)$	δ e K

Fonte: Faceli *et al.*, (2011).

A função Kernel RBF (Radial Basis Function – Função de Base Radial) nos estudos de aprendizado de máquina é utilizada para soluções de problemáticas que envolvem o estudo de máquinas de suporte, justamente pelo ajuste de espaço em alta dimensão, que é um requisito para mapeamento desta função (VIDHYA, 2013).

Em estudos que envolvem regressão, a função do Kernel RBF obtém, em escala de desempenho computacional, uma melhoria superior em relação a seus concorrentes como o

polinomial e o Sigmoidal, isso por ter um número menor de parâmetros e mencionado por muitos estudiosos como primeira escolha nas possibilidades existentes dos Kernels (NOVAKOVIC; VELJOVIC, 2011).

Ademais, destaca-se que um Kernel expressa uma medida de similaridade entre vetores. O Kernel RBF representa esta semelhança como uma função decadente da distância entre os vetores (isto é, a norma quadrada de sua distância). Ou seja, se os dois vetores estiverem próximos junto $\|x - x'\|$ então será pequeno. Então, enquanto $\gamma > 0$, segue-se que $\gamma \|x - x'\|^2$ será maior. Assim, vetores mais próximos têm um valor de Kernel RBF maior que vetores mais distantes. Esta função tem a forma de uma curva em forma de hipérbole. O parâmetro γ define a largura da curva em forma de hipérbole. Quanto maior o valor de γ mais estreita será a hipérbole. Pequenos valores de γ produzem hipérbole largas (SMOLA; SCHÖLKOPF, 2004).

O uso deste Kernel permite a obtenção de separação linear existente entre os espaços dimensionais dos dados de entrada, onde pode ser modificado conforme o ajuste do usuário, com base nos resultados e o uso de métricas regressivas é possível estimar a presença de resultados contínuos e não-contínuos, obtidos através da distância de cada *pixel* mapeado pelo Kernel RBF.

3.3 ÁRVORE DE REGRESSÃO

Segundo Russel e Norwing (2003) a árvore de regressão é uma das formas mais simples e bem sucedidas de algoritmos de aprendizagem. As árvores de regressão pertencem à família do algoritmo CART – *Classification and Regression Trees*, com uma metodologia que trabalha com previsão ou regressão de modelos de desempenho matemático e computacional. Tais algoritmos utilizam uma abordagem com configuração dividir para conquistar, buscando dividir em vários subconjuntos os exemplos utilizados para o treino, até cada um destes subconjuntos pertencer a uma mesmo escalar, ou até sair escalares majoritária (QUINLAN, 1993).

As árvores de regressão são principalmente de uso, para fazer explicações e fazer previsão para determinados atributos. Tais atributos são valores observados para explicar mudanças e não-mudanças num cenário onde o modelo gerado pelo algoritmo resulta com características do mesmo atributo permitindo a construção homogênea desses atributos (nós da árvore de regressão). Por ser um uma árvore binária, ressalta-se a facilidade de interpretação e leitura por seus usuários.

Uma árvore de regressão pode ser definida recursivamente como:

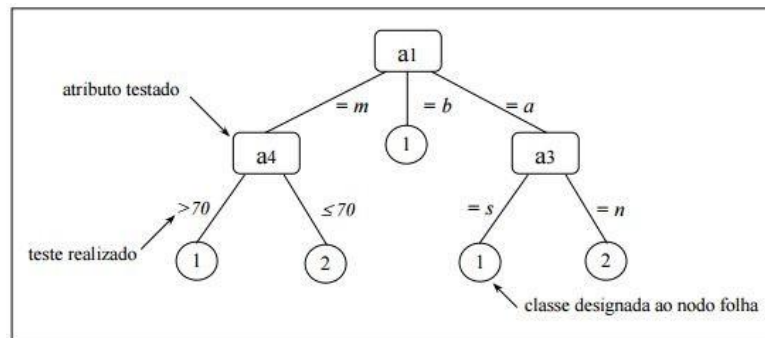
- Um nó folha que corresponde a um escalar;

- Um nó de regressão, que contém um teste sobre algum atributo. Para cada resultado do teste existe uma aresta para uma sub-árvore.

A Figura 07 apresenta a representação de uma árvore de decisão, no qual são testados atributos quantitativos e categóricos.

A árvore de regressão tem como entradas no nosso caso de estudo os *pixels* do histórico das imagens da série temporal, resultando numa saída do pixel no tempo futuro. O procedimento é repetido para cada *pixel*, até varrer as imagens completamente. Além disso, existe uma forte compressão das sequências de imagens, o que reduz redundâncias e concentra a maior parte da informação dentro de poucos coeficientes.

Figura 07 – Exemplo de uma árvore de Regressão.

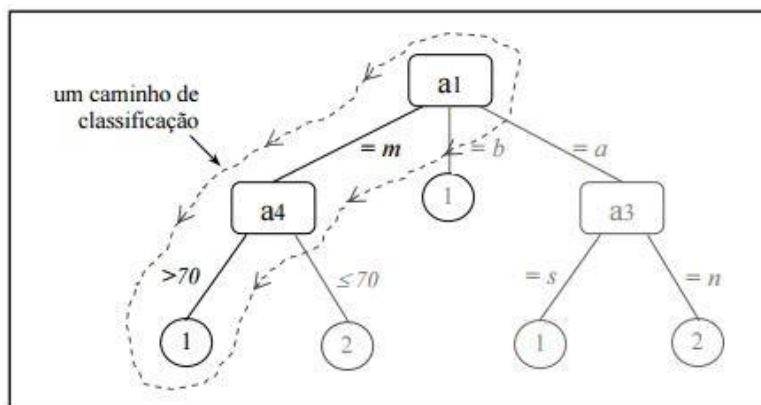


Fonte: Araújo (2016).

Neste exemplo da Figura 07, a1, a3 e a4 representam os atribuídos distribuídos na árvore de acordo com seu nível informativo. Nos arcos, são testados os valores os atributos designados ao nó a que pertencem. Valores categóricos, são representados por uma igualdade, por exemplo $= m$, onde m é um valor do atributo. Para valores quantitativos e entre o intervalo do valor, por exemplo, >70 , sendo este intervalo obtido através cálculo. Cada círculo, ao final dos ramos da árvore, indica a previsão associada aos nós das folhas. A regressão ocorre ao se percorrer o caminho que se inicia no nodo raiz ($a1$) e se estende até as folhas. Para a regressão, os valores das folhas são quantitativos.

A regressão ocorre ao se percorrer o caminho que se inicia no nodo raiz ($a1$) e se estende até as folhas, Figura 08. Para a regressão, os valores das folhas são quantitativos.

Figura 08 – Exemplo de um caminho percorrido.



Fonte: Araújo (2016).

O caminho destacado na Figura 08 pode apontar uma regressão. Através do caminho descrito por uma árvore de decisão é possível derivar regras. As árvores e as regras são geralmente utilizadas em conjunto. De acordo com a aplicação, quando as árvores tendem a crescer muito, estas são substituídas pelas regras.

A escolha do melhor atributo ocorre de forma a utilizar aquele que tenha melhor capacidade isolada de classificar os exemplos. A maioria dos algoritmos de indução da árvore trabalha com uma função de divisão univariável, onde cada nó interno é dividido de acordo com um único atributo. O algoritmo busca o melhor atributo para ser o nó-pai, da árvore ou sub-árvore, de maneira que minimize ao máximo o grau de impureza dos nós filhos (ZUBEN; ATTUX, 2008).

A árvore de regressão é uma ferramenta de decisão cuja estrutura treinada pode revelar um conjunto de regras explanáveis e com certa precisão. O problema de regressão consiste na criação de um modelo funcional que relacione o verdadeiro valor da saída com os regressores (variáveis de entrada). Aplicar esta técnica, assume-se que a variável de saída era emular uma previsão, utilizando-se para a resolução de problemas de regressão. O processo de treino de uma árvore de regressão consiste num algoritmo de partição do conjunto de treino, em cada nó da árvore, em dois subconjuntos disjuntos (tratando-se, então, de uma árvore binária), assim a árvore é criada com base em regras de parada que interferem e contribuem na melhoria da previsão pela árvore (MERINO; NUNES, 2011).

Nas regras de divisão existe um conceito fundamental chamado de ganho de informações, o qual utiliza o conceito de entropia, que tem como base medir a aleatoriedade de uma variável aleatória. A Árvore de regressão usa a entropia para medir a aleatoriedade (complexidade em fazer previsões) na variável alvo (meta). Todos os nós de regressão, o atributo que mais reduz a aleatoriedade da variável meta é nomeado para dividir os dados. O

ganho de informações mede a redução da entropia para cada atributo, a partir das partições adquiridos de acordo com os valores do atributo. A construção de uma árvore de regressão é guiada pelo propósito de minimizar a entropia, que refere-se a complexidade em fazer predições da variável alvo (meta) (NORVIG; RUSSEL, 2014). A entropia de uma variável aleatória é atribuído com v valores que são valores em propósito de minimizar a entropia $v = [v_1, v_2, v_3, \dots, v_k]$, a métrica da entropia é em bits $B_{\frac{p_k}{(p_k+n_k)}}$ complementando as informações a determinada pergunta, e utilizando o logaritmo na base 2, para cada um com probabilidade $P(v_k)$, é definida pela equação:

$$H(v) = \sum_k P(v_k) \log_2 \frac{1}{P(v_k)} = - \sum_k P(v_k) \log_2 P(v_k) \quad (6)$$

O ganho de informações dos testes aplicados no atributo pode ser visto na equação que representa a redução prevista na entropia:

$$\text{Ganho} = B\left(\frac{p}{p+n}\right) - \sum_{k=1}^d \frac{p_k+n_k}{p+n} B\left(\frac{p_k}{p_k+n_k}\right) \quad (7)$$

Onde p representa os exemplos positivos e n exemplos negativos de entrada. Para cada subconjunto tem p_k que representa os exemplos positivos e n_k os exemplos negativos de entrada. Se continua ao longo dessa ramificação, precisamos de $B(p_k/(p_k+n_k))$ bits para complementar as informações para responder à interpretação do problema.

No problema de regressão, a função de custo que reduz é, geralmente, o erro quadrático. Mediante disso, a constante relacionada às folhas de uma árvore de regressão é a média do atributo alvo (previsores) das amostras de treinamento que caem da folha. Na elaboração de uma árvore de regressão, ela tem como base a árvore de decisão devido a sua recursividade e assim é praticamente tudo semelhante ao seu desenvolvimento, o que diferencia a árvore de regressão para a de decisão é a função de custo ou seja, enquanto a árvore de regressão minimiza os custos, a árvore de decisão maximizar esses custos na otimização é uma parte da matemática que inclui características de resultados para formar funções que devem ter suas restrições e que podem ser impostas (FACELI *et al.*, 2011; NORVIG; RUSSEL, 2014).

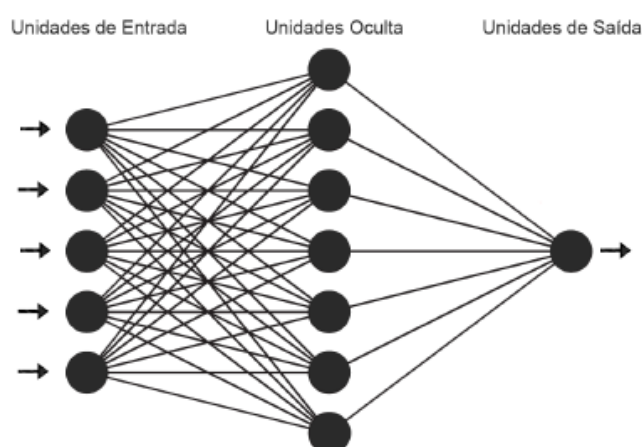
3.4 REDE NEURAL ARTIFICIAL

Redes neurais (RNA) são, provavelmente, a mais antiga técnica de inteligência artificial em uso, tendo sido desenvolvida na década de 1940. Uma RNA típica é constituída por um conjunto de (N) neurônios interligados, influenciando uns aos outros e formando um sistema maior capaz de armazenar conhecimento adquirido, principalmente por meio de exemplos pré-

apresentados e assim podendo realizar inferências sobre novos exemplos, onde são apresentadas situações desconhecidas (HAYKIN, 2001).

As redes neurais *Multi Layer Perceptron* (MLP) normalmente são apresentadas como um grafo, onde os vértices são os neurônios e as arestas as sinapses, informando a direção das arestas o fluxo das informações (sinais de entrada). O presente modelo é ilustrado na Figura 09.

Figura 09 – Representação de uma Rede de Neurônios Artificiais.



Fonte: adaptado Faceli *et al.* (2011).

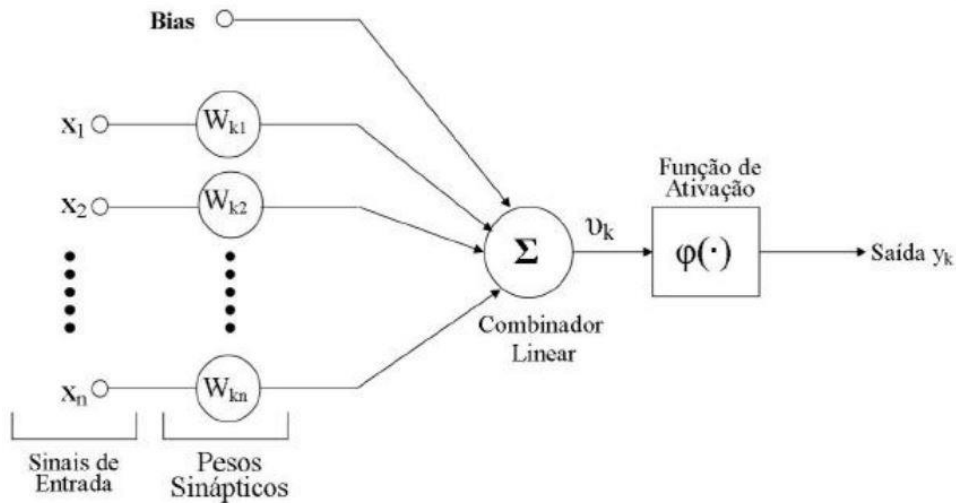
Como citado, os nós, representam os neurônios, unidades básicas do cérebro humano que tem por função transmitir informações, estando introduzidas nestas propriedades de excitabilidade e condução de mensagens nervosas.

As sinapses, representadas pelas arestas do grafo, são as transmissões das informações entre os neurônios. O fluxo das informações, representado pela direção setas, é a discretização do modelo biológico em um modelo matemático, onde existe uma hierarquia de camadas de neurônios. Atualmente, existem os mais diversos tipos de modelos de redes neurais, cada um com especializações e modelo de aprendizado distintos. Segundo Santos (2008), a escolha da melhor configuração de uma RNA para solucionar um problema específico envolve a definição dos seguintes parâmetros:

- Topologia de rede – Conectividade, tipos de conexões, ordem de conexões e intervalos dos pesos;
- Propriedades dos neurônios – Intervalo de ativação e função de transferência;
- Dinâmica do sistema – Início dos pesos e regra de aprendizagem.

O modelo básico, representado pelo diagrama de blocos da Figura 10, mostra a relação entre os elementos do neurônio artificial:

Figura 10. Modelo de um neurônio.



Fonte: Adaptado de Faceli *et al.* (2011).

O neurônio tem os seguintes elementos:

- 1) sinais de entrada x_1, x_2, \dots, x_n .
- 2) um conjunto de sinapses com os seus pesos sinápticos $w_{k1}, w_{k2}, \dots, w_{kn}$. Os pesos sinápticos são representados por w_{kj} , onde k representa o índice do neurônio e j refere-se ao terminal de entrada da sinapse ao qual o peso sináptico se refere.
- 3) Um combinador linear u , representado pela soma dos sinais de entrada ponderado por pesos sinápticos;
- 4) A função de ativação φ para restringir a extensão da saída de um neurônio;
- 5) Um bias é a hipótese ou dados que têm o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo, respectivamente.
- 6) v_k é a saída gerada depois dos dados entrarem no *combinador linear*.

O algoritmo de treinamento utilizado com a rede neural MLP para a previsão de séries temporais foi o Levenberg-Marquardt. Este algoritmo tem sido amplamente utilizado na literatura em problemas gerais que envolvam redes neurais, incluindo redes recorrentes, e se encaixa no paradigma de aprendizagem supervisionada, pois utiliza dados conhecidos como treinamento para prever resultados futuros. O algoritmo de Levenberg-Marquardt é baseado no método de otimização de Newton para sistemas não-lineares. Na prática, o método de

otimização de Newton para treinamento não é utilizado diretamente em função da grande complexidade, do elevado tempo computacional, ou mesmo impossibilidade da estimação das derivadas de segunda ordem (HAYKIN, 2001).

$$W(n+1) = W(n) - (J^t(n) J(n) + \mu I)^{-1} J^T(n) e(W) \quad (8)$$

Na equação, J é a matriz jacobiana das derivadas dos erros relacionados aos pesos, μ é um valor escalar e (W) é um vetor de erros. Se o valor de, μ é grande, a expressão acima representa uma descida mais rápida, para pequenos valores de μ a expressão é reduzida para o método de Gauss Newton, que é mais rápido e mais preciso, próximo do erro mínimo. No caso de estudo, o treinamento foi feito com o ajuste dos parâmetros livres da rede, adaptando-a gradativamente ao ambiente externo.

A estimação e previsão de séries temporais de imagens (2D), é um processo que depende de um grande conjunto de probabilidades, onde regiões (conglomerados de pixels) podem crescer e se deslocar para quaisquer direções (AKBAR *et. Al.*, 2012). A seguir, serão propostas estratégias de previsão de sequências de imagens sem que isso acarrete num problema de dimensionalidade.

A previsão no domínio pixel é realizada quando cada *pixel* da imagem da série temporal está como uma função da sua intensidade de brilho, de modo que a mudança de intensidade de *pixel* é ajustada pela função de saída da rede neural. Este é o método de previsão direta, sem qualquer transformação matemática.

3.5 PREVISÃO E MUDANÇAS

No meio computacional são realizados testes para análise do desempenho dos algoritmos preditores utilizados, com o intuito de se observar mudanças da série temporal e, consequentemente, objetiva a execução de preditores que realizem a previsão do próximo valor da serie original. Com a janela (resultados) mais semelhante à série temporal original, que é obtida com o uso dos preditores, é selecionado o melhor preditor f (RIBEIRO *et al.*, 2011). Tal seleção é feita pelo cálculo da distância entre seus eixos, onde se consegue fazer comparações e mudanças que possam ocorrer desde o treinamento até a saída dos resultados. Desta forma, a escolha do melhor preditor da série temporal dá-se tanto pela saída dos valores preditores quanto dos padrões de comparação de comportamento da série original em estudo (AMORIM NETO *et al.*, 2010).

4 ÁREA ESCOLHIDA PARA O ESTUDO

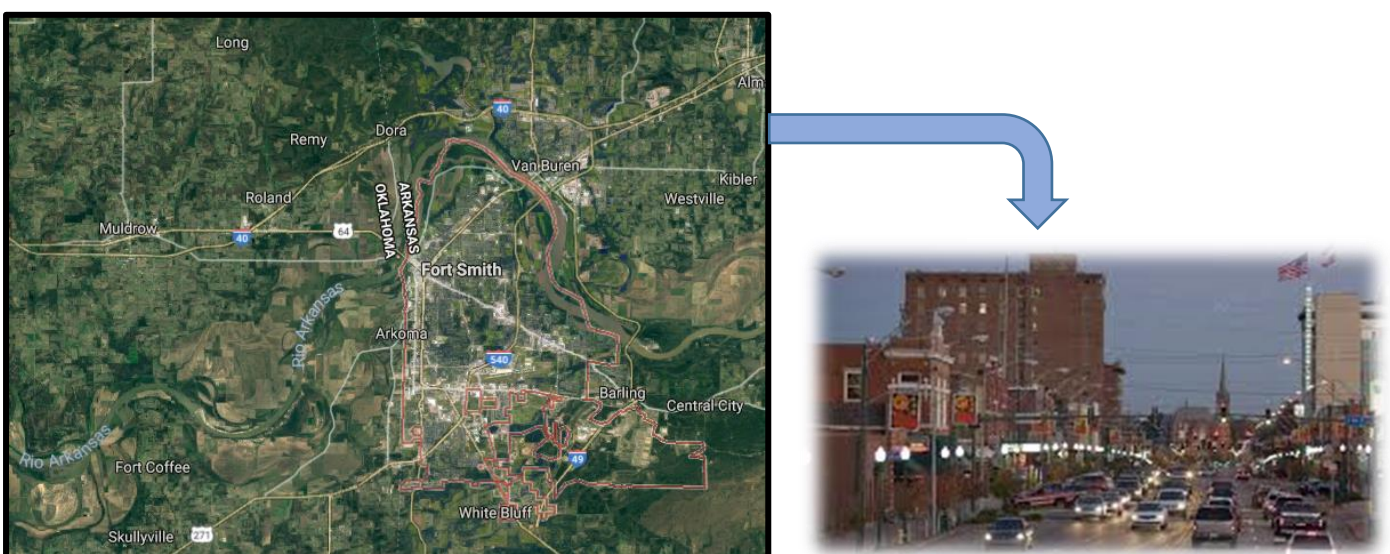
A área de estudo escolhida mostra a variação dos recursos hídricos para serem cortados e processados foi correspondente à região localizada na cidade de Fort Smith, que é cortada pelo Arkansas River. Assim, será apresentada uma breve sinopse da cidade e do rio que é o foco do caso de estudo. A escolha desse local deu-se ao fato da similaridade das condições climáticas terem aparência com a da região nordeste, fora que o período de não aparecimento de chuvas é de 8 a 9 meses.

E por ser uma nova abertura para utilização do algoritmo SVR no que se diz respeito ao seu desempenho em relação aos algoritmos Árvore de Regressão e as Redes Neurais.

4.1 Cidade de Fort Smith

A cidade de Fort Smith, situada no estado americano do Arkansas, nos Estados Unidos, tem uma área de 170,5 km², dos quais 160,5 km² estão cobertos por terra e 10 km² cobertos por água. De acordo com dados populacionais de 2010, a população da cidade é de 86.209 habitantes, e sua densidade populacional é de 537,12 hab/km². É a segunda cidade mais populosa do Arkansas. A cidade possui 37.899 residências, a Figura 11 retrata a cidade de Fort Smith vista pelo satélite e da sua entrada (GCT- PH1, 2011).

Figura 11- Região e entrada da cidade de Fort Smith.



Fonte: Gct- ph1 (2011).

4.2 Rio Arkansas

O rio Arkansas nasce nas montanhas rochosas e faz seu percurso pelos estados do Colorado, Kansas, Oklahoma e Arkansas. Antes, ele servia de limite de fronteira com o país do México, como é citado no Tratado de Adams-Onís entre o México e os Estados Unidos, em 1819. É o sexto mais longo rio dos Estados Unidos e o maior tributário do sistema Mississippi-Missouri (MCCLELLAN, 2010).

O rio Arkansas tem origem perto de Leadville e desagua no rio Mississippi, na cidade de Napoleon, equidistante de Memphis a montante e Vicksburg, a jusante. É o maior afluente da bacia inteira do sistema fluvial Mississippi-Missouri, em termos de área drenada (505.000 km²), a Figura 12 mostra o recorte para estudo do rio Arkansas e a vista do Rio Arkansas perto da cidade (MCCLELLAN, 2010).

Figura 12- Rio Arkansas em local recortado para estudo e visualizado em Fort Smith.



Fontes: Autoria Própria; Mcclellan (2010).

Essa área, por conter um certo número populacional e algumas fábricas, foi escolhida para mostrar que o estudo feito com predições pode fazer uma análise de como a ação do homem ou da própria natureza pode afetar o rio, de modo a viabilizar a resolução dos problemas hídricos em locais ínfimos com maior celeridade pelas autoridades competentes, trazendo mais agilidade na tomada de decisões no que concerne à diminuição/aumento de danos a esse recurso natural.

A utilização destas imagens foi dada também pela parceria com pesquisadores do Centro de Pesquisas Espaciais da University of Texas, Austin, pela similaridade climática e pelo tempo de prolongado de aparecimentos de chuvas no local.

4.3 Desenvolvimento e importância dos rios

Os recursos naturais são de extrema importância quando se trata de desenvolvimento industrial, como principal componente dos rios a água que é um recurso finito e de fator concorrência no mercado.

Ao longo da história de crescimento de industriais nas nações o “local de instalação” foi um fator determinante que pode colaborar no desenvolvimento dessas indústrias que é a proximidade a rios e lagos. Entretanto, no perspectivado mercado global, hoje torna-se mais importante como utilizar a água disponível em rios, lagos e subterrâneos do que ficar ostentando sua abundância (MARTINEZ, 1991)

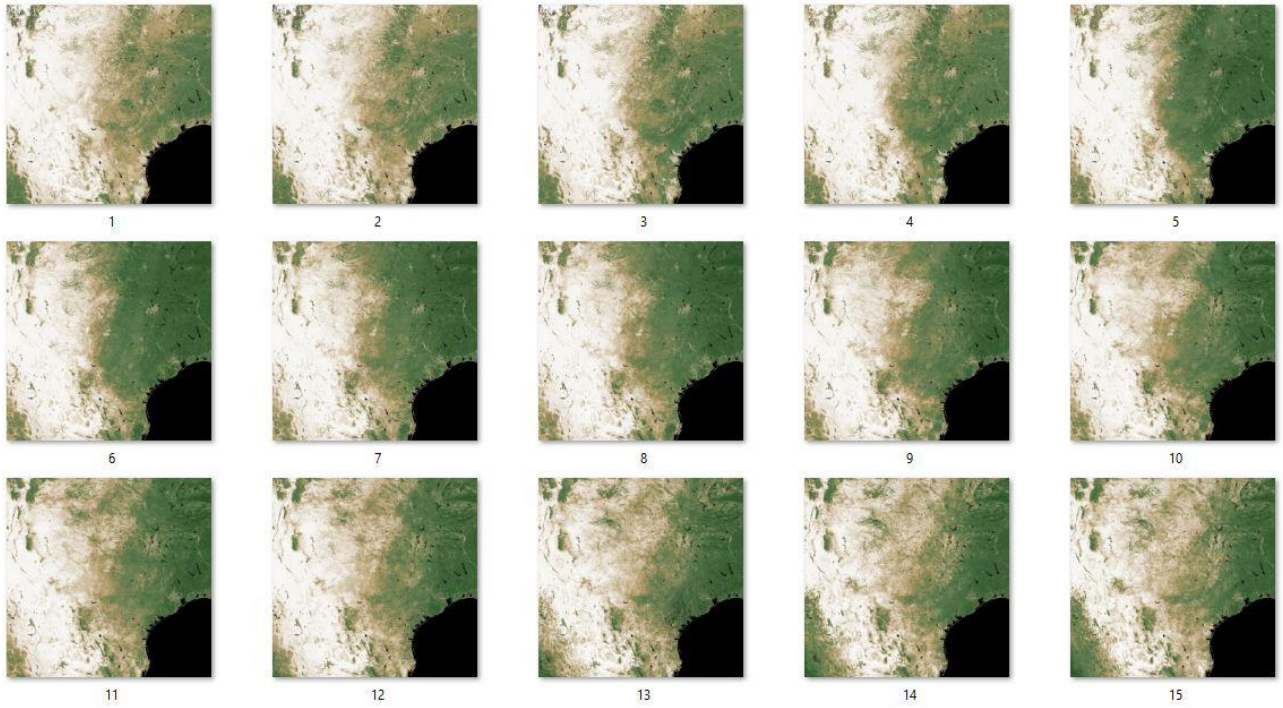
Crescimentos da produção mundial com superação no acréscimo populacional, vêm tirando a tranquilidade de muitas pessoas com relação ao uso da água, ao lado alimentos que não tem a mesma qualidade e a degradação dos solo com baixa resposta produtiva, (CHRISTOFIDIS, 1997), se continuar desta forma sem uma ponte de equilíbrio, a perda da estabilidade econômica terá seria consequências, como crises sociais de difícil solução.

Assim se não houver uma acentuada diminuição dos grandes desperdícios no uso desses recursos, práticas desleais serão utilizadas prejudicando cidadãos de grandes e pequenas cidades fora o meio rural.

4.4 Sequência de Imagens de satélite

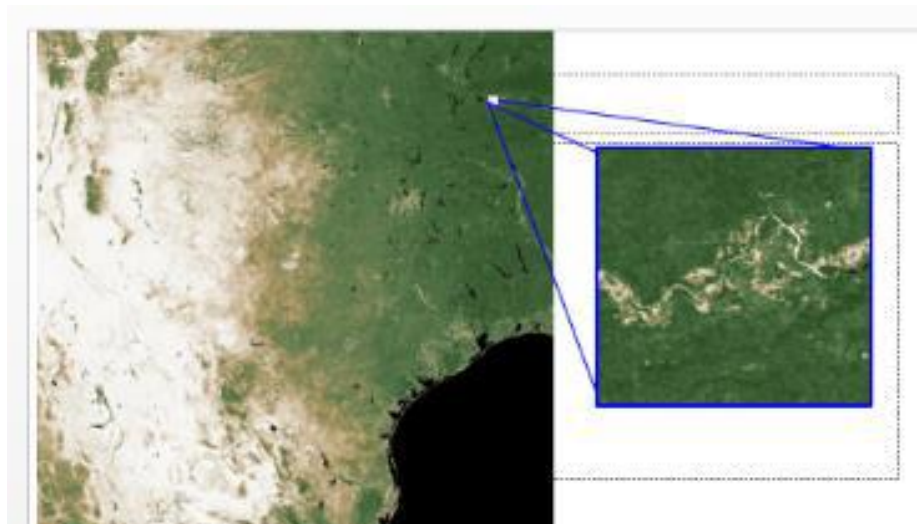
Devido às dimensões muito grandes das imagens brutas originais, foi necessário escolher uma área de estudo que mostrasse a variação dos recursos de água para ser cortada e processada. A área escolhida foi correspondente à cidade de Fort Smith, que é cortada pelo rio Arkansas, o satélite utilizando foi o AQUA, com o sensor MODIS. Datadas do ano de 2002, pelo satélite, essas imagens foram fornecidas pelo Centro de Pesquisas Espaciais da Universidade do Texas, em Austin como mostra as figuras 13 e 14 e 15.

Figura 13: Série temporal de imagens da área do Texas e regiões envolventes.



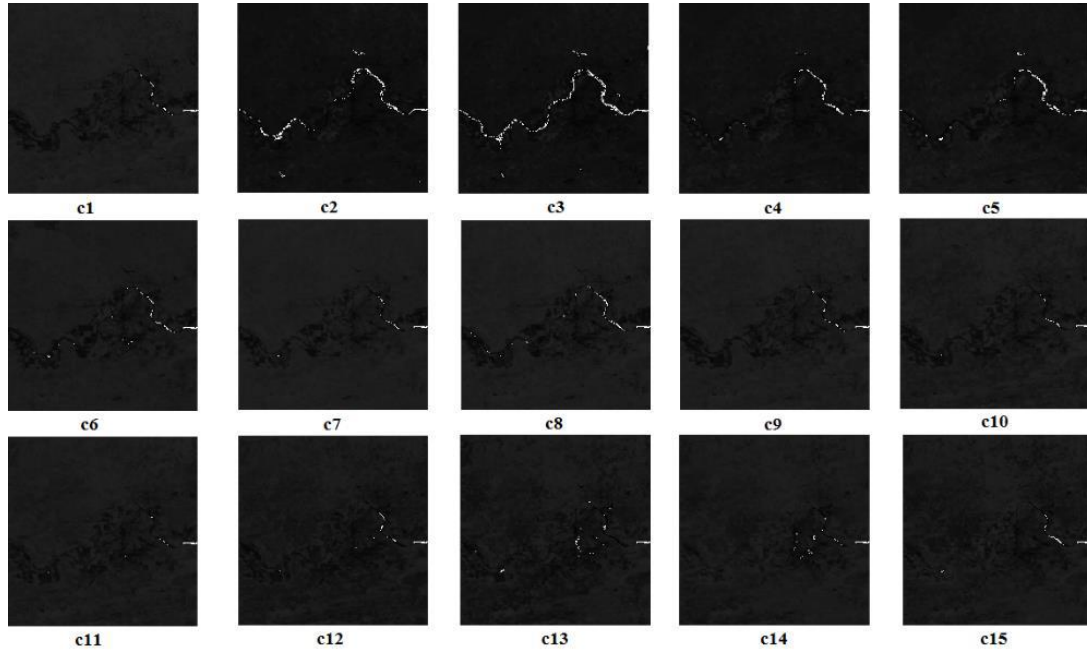
Fonte: MODIS (2016).

Figura 14 - Série temporal de imagens da área do Texas e regiões envolventes.



Fonte: MODIS (2016).

Figura 15: Série Temporal de imagens recortadas do Rio Arkansas.



Fonte: MODIS (2016).

5 ANÁLISE DAS IMAGENS POR MEIO DE METRICAS GLOBAIS REGRESSIVAS

Foram utilizadas 15 imagens do MODIS processadas em NDVI Figura 15, para se fazer o treinamento. A partir dessa série, foi realizada a predição utilizando os algoritmos de forma individual (Rede Neural, Árvore de Regressão e o SVR) da seguinte forma:

- Predição da 15ª imagem, usando-se as primeiras 14 imagens em NDVI para treinamento;
- Predição da 14ª imagem, usando-se as primeiras 13 imagens em NDVI para treinamento;
- Predição da 13ª imagem, usando-se as primeiras 12 imagens em NDVI para treinamento;
- Predição da 12ª imagem, usando-se as primeiras 11 imagens em NDVI para treinamento;
- Predição da 11ª imagem, usando-se as primeiras 10 imagens em NDVI para treinamento.

Nos algoritmos, são aplicadas imagens do MODIS processadas em NDVI conforme a classificação feita no domínio de espaço (*pixel a pixel*), onde essa técnica tem predição direta. Foram utilizados dados entradas com 150x150 de dimensão dos *pixels* das imagens da série temporal, na matriz $I_l(i, j)$ a $I_n(i, j)$ (FACELI *et al.*, 2011), resultando numa saída de dados (*pixel*) de tempo futuro $I_{n+1}(i, j)$. O mesmo processo é repetido para cada *pixel* das imagens em análise para cada algoritmo utilizado na pesquisa.

Essa metodologia foi utilizada para as imagens sintéticas que foram utilizadas para validação dos algoritmos.

5.1 METODOLOGIA GLOBAL REGRESSIVA PARA VALIDAÇÃO

A metodologia de análise de um algoritmo de aprendizado de máquina como exemplo do modo supervisionado é normalmente concretizado, realizado através de uma averiguação do desempenho do atributo de predição por ele na assimilação de novas características ou novos atributos, não mostradas anteriormente em seu treinamento (FACELI *et al.*, 2011).

Comumente, são utilizadas duas métricas para se fazer a predição, são elas: *Medium Absolute Error* (MAE) e o *root-mean-square-error* (RMSE). A justificativa para utilização

dessas medidas dá-se pela necessidade de se medir a qualidade dos modelos e também o poder de previsibilidade do modelo. Um erro é definido como a diferença entre o valor real e o valor previsto; portanto, quanto menores são os erros, melhor é o modelo e a capacidade de prever valores futuros (LIMA, 2009; ADAMICK; BRÉDA; VAISPACHER, 2015).

O MAE e o RMSE podem ser usados em conjunto para diagnosticar a variação nos erros em um conjunto de previsões. No nosso caso de estudo, este foi aplicado sobre os pixels das imagens real e estimada, segundo a equação (9). O MAE é uma métrica de erro estatístico usado para medir a grandeza da proximidade entre o resultado previsto e o real para expressar a confiabilidade dos modelos. Quanto mais próximo o valor da métrica de zero, mais eficaz é o método de previsão (LIMA, 2009; DE MYTTNAERE, 2016). Nas equações seguintes, x_i corresponde ao valor real, y_i corresponde ao valor previsto e “ n ” é o número de repetições.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (9)$$

Já o RMSE ou raiz quadrada média (ou, por vezes, root-mean square de erro) é uma medida utilizada frequentemente das diferenças entre os valores (amostra) previstos por um modelo ou um estimador e os valores observados. O RMSE representa o desvio Médio da amostra das diferenças entre os valores previstos e os valores observados (FACELI *et al.*, 2011).

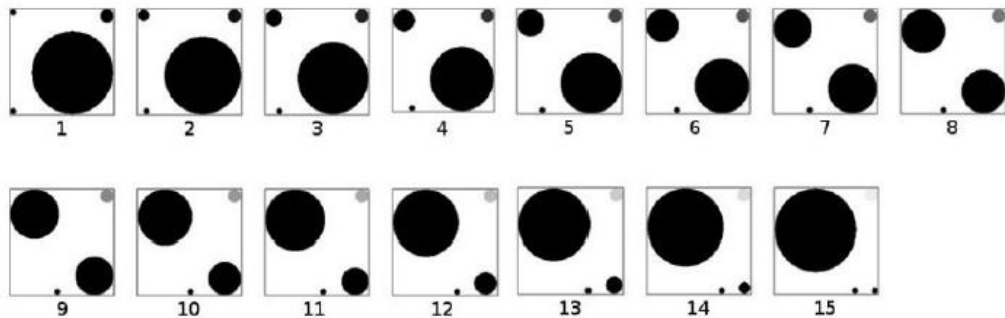
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (10)$$

No trabalho de Teixeira *et al.*, (2014) são utilizadas ambas as métricas para se avaliar o desempenho dos algoritmos de regressão.

5.2 IMAGENS SINTÉTICAS

Para validar, também, o uso dos algoritmos de aprendizado de máquina foram utilizadas 15 (quinze) imagens sintéticas usando o método que permite a predição de variação dos níveis de cinza, movimentação, aumento e diminuição de locais da séries temporais de imagens. Os testes dessas imagens sintéticas têm sua importância na validação dos algoritmos (RADKE *et al.*, 2005) e, por essa razão, foram produzidas 15 imagens com dimensões de 150x150 *pixels*, com formas geométricas e as seguintes mudanças: aumento, diminuição, movimento e modificação nos níveis de cinza, todos ao mesmo tempo na mesma série temporal, conforme representado na Figura 16 abaixo.

Figura 16: Série temporal de imagens sintéticas com as seguintes figuras geométricas.



Fonte: Autoria Própria.

Nestas imagens, foram utilizadas as predições do 13^o ao 15^o e foram calculados o erro MAE e RMSE.

5.3 METODOLOGIA LOCAL PARA VALIDAÇÃO

Com o intuito de acelerar e tornar o processo de estimativa de séries temporais de imagens, foram produzidos os mapas temáticos dos locais onde o rio mostrou mudanças ao longo do tempo, onde o *plugin* do Qgis chamado de *Semi-Automatic Classification Plugin* (SCP) foi usado nos procedimentos de classificação das imagens em NDVI do rio Arkansas (SUTTON, 2012).

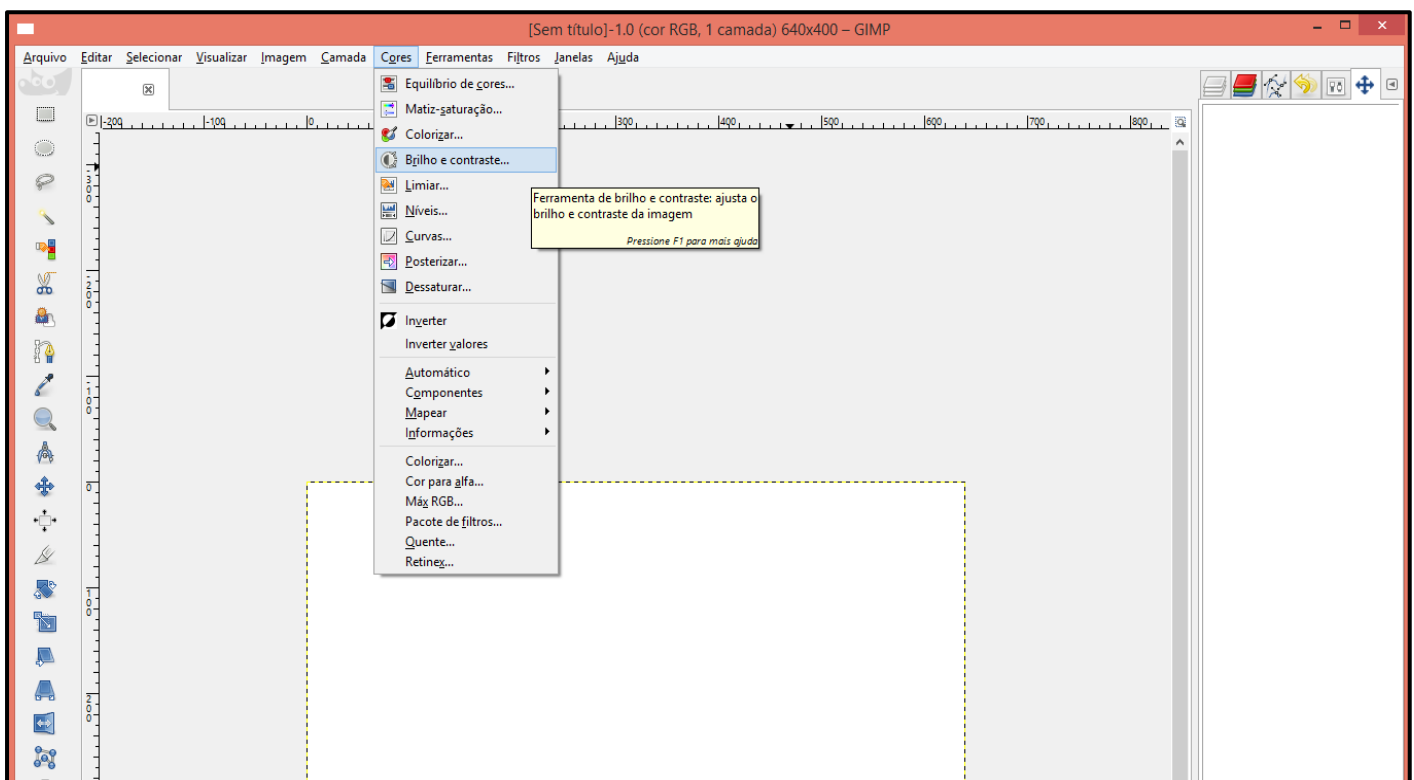
O *Plugin* de Classificação Semiautomático (SCP) permite a classificação supervisionada de imagens de sensoriamento remoto, fornecendo ferramentas para o *download*, o pré-processamento e o pós-processamento de imagens (SUTTON, 2012). O desenvolvimento desta ferramenta foi feito por Lucas Gongedon, um dos desenvolvedores das ferramentas de utilidade do Qgis, onde esse *software* livre permite desenvolver mapas com variações de camadas usando variedade de projeções de mapa. Mapas podem ser confeccionados em diferentes formatos e para diferentes finalidades. Qgis desenvolvido por *Qgis Development Team*, permite desenvolver a representação de mapas a partir de vetoriais. A variedade e os diferentes tipos de imagens vetoriais são suportados e o *software* tem capacidade de georreferenciamento das imagens (SUTTON, 2012).

Nos passos iniciais, foi feita a segmentação do rio nas demais faces (vegetação densa, vegetação rasa, solo exposto), onde o procedimento de binarização das imagens foi executado para realçar as feições do rio. Essa binarização foi feita com o *software*–livre Gimp, que contém

a ferramenta na aba *cores* para melhoria do contraste e visualização da face do rio. Assim, foi possibilitado um realce nas mudanças ao longo do tempo.

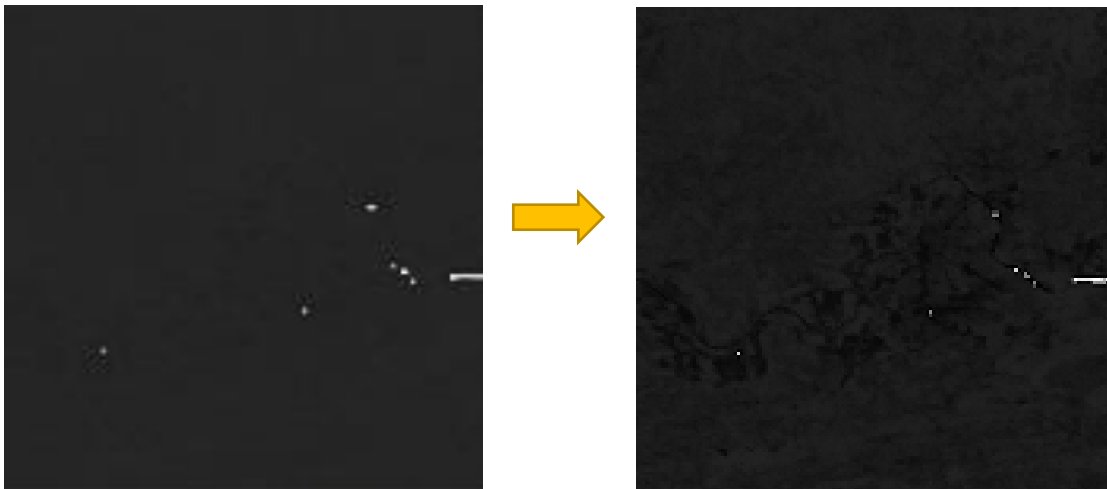
Na Figura 17, será mostrada a interface de cada *software* para melhor entendimento do processo de binarização a classificação. A figura abaixo mostra a interface do *software* Gimp com o principal comando para binarizar a imagem, já na figura 18 mostra o antes e depois do processo de binarização:

Figura 17- Interface do *Gimp*..



Fonte: Autoria Própria.

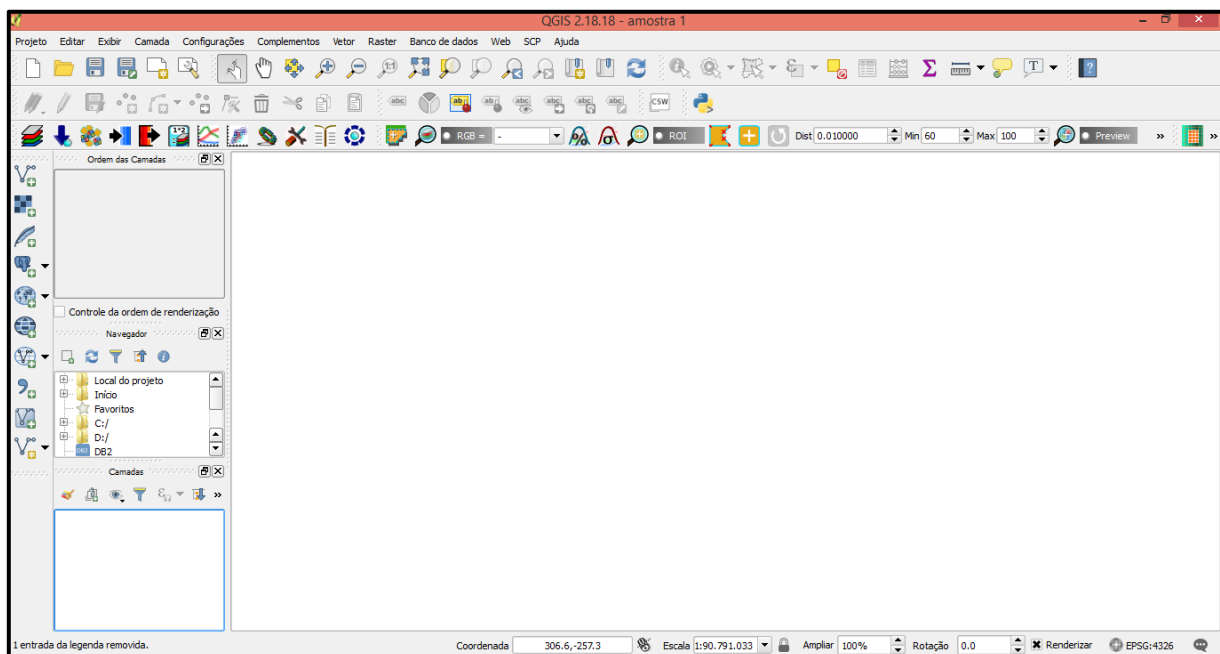
Figura 18: Antes e depois da binarização com realce no rio Arkansas.



Fonte: Autoria Própria.

Depois de feita a binarização começa a ser feito o processo de classificação, onde foi utilizado o *software* Qgis 2.18.18, a seguir a Figura 19 mostra a interface do *software*:

Figura 19: Interface do Qgis 2.18.18.

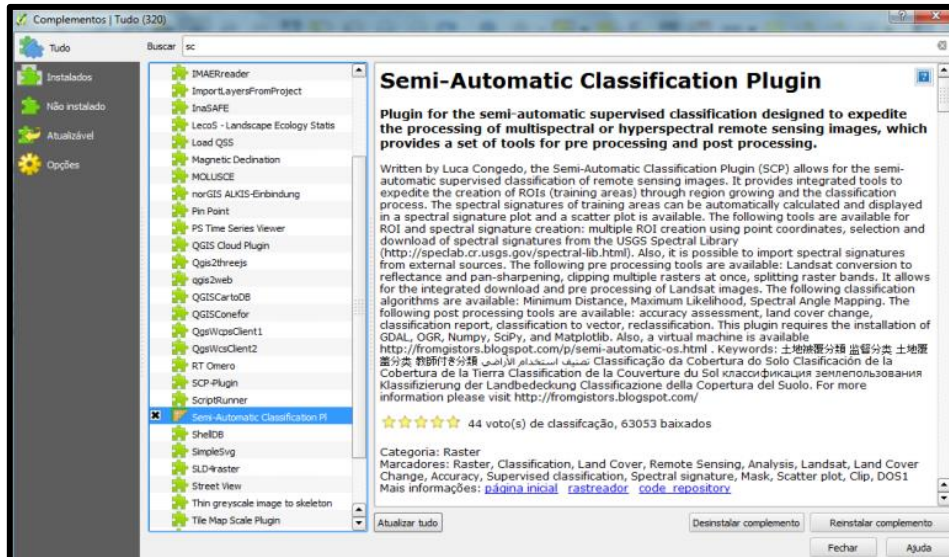


Fonte: Autoria Própria.

Com a interface aberta foi escolhida uma imagem *raster vetorial*, que pode ser nos formatos (PNG, JPEG, TIFF, GEOTIFF, SHAPE). Com a imagem escolhida, e na área de trabalho, procura-se pelo complemento SCP, *o semi-automatic classification plugin (SCP)* que

é um complemento desenvolvido para o *software* Qgis que possibilita a realização de classificação *pixel a pixel*, o *semi-automática* ou supervisionada de imagens, figura 20 capturadas por diversos sensores e satélites (SUTTON, 2012).

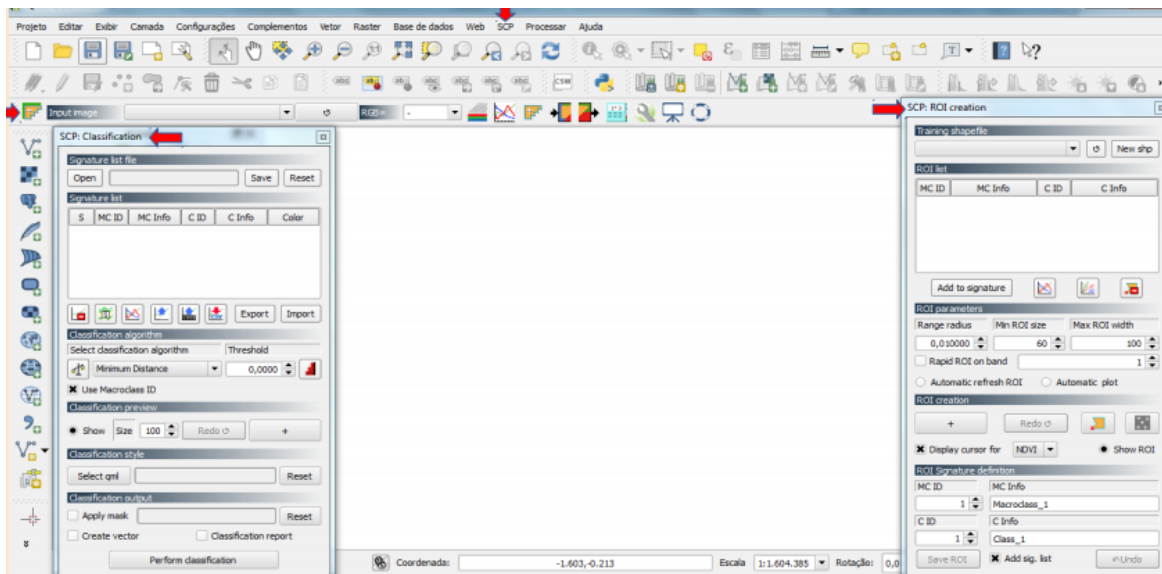
Figura 20- Complementos – Gerenciador e instalação de complementos do *Plugin* (SCP).



Fonte: Autoria Própria.

Após a instalação do complemento, suas ferramentas ficam disponíveis através do menu, da barra de ferramentas e de dois painéis como mostra a figura 21.

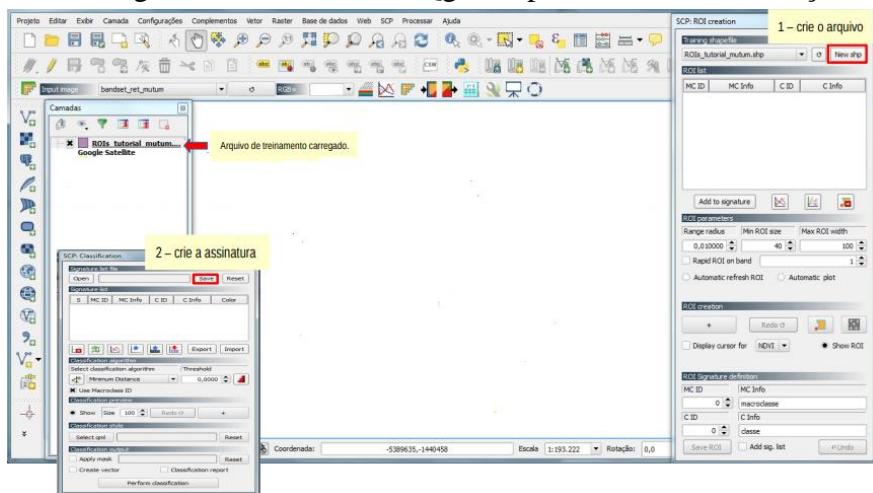
Figura 21- Obtenção de Imagem.



Fonte: Autoria Própria.

A primeira etapa da classificação consiste em definir categorias de interesse por meio da seleção de amostras onde se faz a coleta conforme a tonalidade da cor daquele *pixel*. Primeiramente, habilita-se o painel SCP: *ROI creation* para a criação do arquivo de treinamento. Logo após, abre-se o painel SCP: *Classification* e salva-se o arquivo que conterà as *assinaturas espectrais*. Na Figura 22 é mostrado o retorno ao painel *ROI Creation* para criar alvos vinculados à *Macroclasses* e *Classes*. Imagens de resolução um pouco melhor podem ser auxiliares neste processo (SUTTON, 2012).

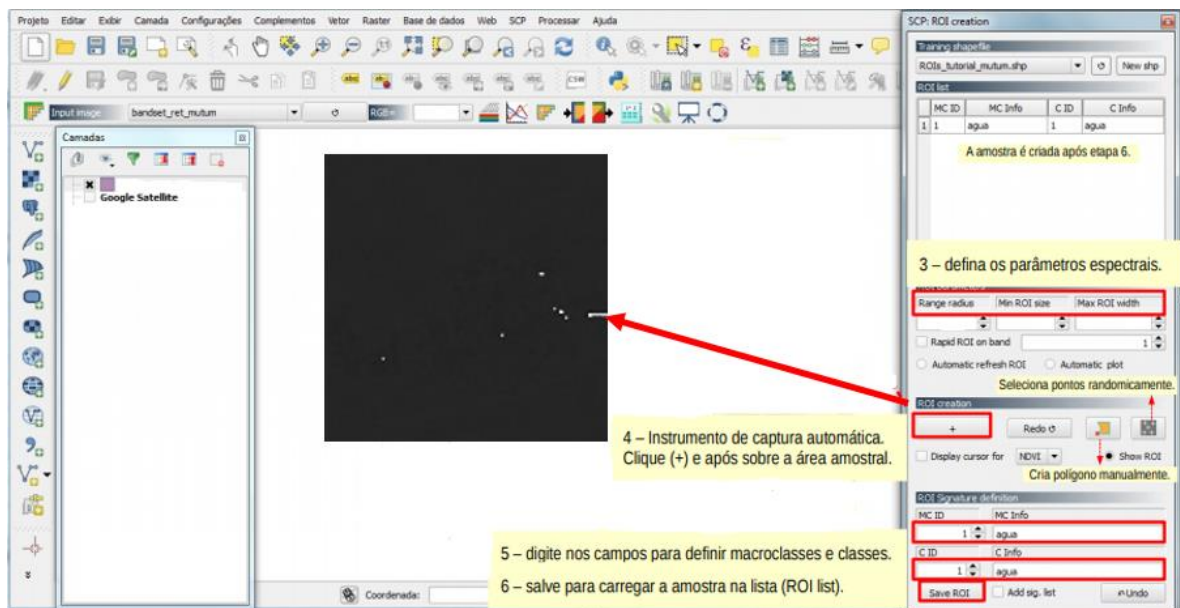
Figura 22 - Interface do Qgis no processo de classificação.



. Fonte: Autoria Própria.

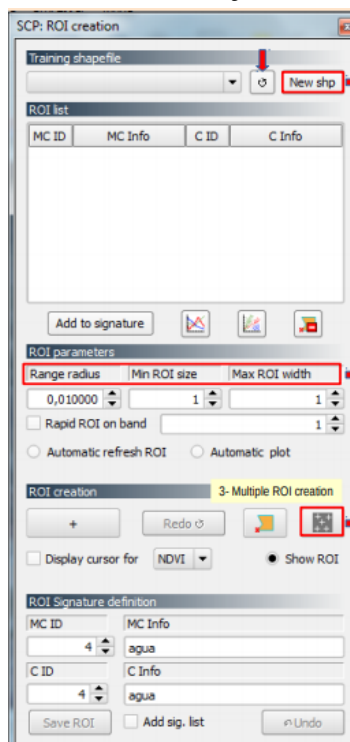
Nas imagens 23 e 24 é mostrado no caso em estudo, como as imagens foram binarizadas, foram criadas duas classes: a classe (rio) e a classe (solo exposto, vegetação densa, vegetação rasa).

Figura 23: Inserção e Processo de Classificação.



Fonte: Autoria Própria.

Figura 24: Processo de seleção de amostras aleatórias com fins de estimativa da acurácia da classificação.



Fonte: Autoria Própria.

Selecionando amostras aleatórias com fins de estimativa da acurácia da classificação, esta fase faz parte do pós-processamento e inicia com a criação do *shapefile*, que conterà amostras selecionadas aleatoriamente por meio da ferramenta *Multiple ROI creation* (SUTTON, 2012).

Depois de feita a classificação é realizado o processo de visualização das áreas escolhidas. Nesse processo é estabelecido o quantitativo de classes por área ou por contagem de *pixel* pelo Qgis. Com o algoritmo *r.report*, é possível obter resultados de área para os arquivos *raster* e aplicar essa técnica no mapeamento de áreas classificadas (SUTTON, 2012).

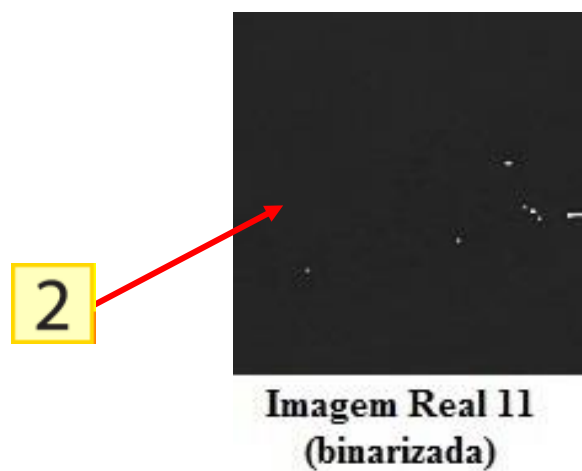
Na nossa imagem *raster*, mostrado na Figura 23 para as classes (Rio) e (Solo Exposto, Mata Densa, Mata Rasa), é possível visualizar o valor do *pixel* na imagem. Usa-se a ferramenta *Identificar Feições* (1) Figura 25 e clica-se sobre uma área qualquer (2) do *raster*, Figura 26. No comando *Tabela* (3), e (4) será o valor exposto, figura 27 visualiza-se o valor do *pixel* da área classificada.

Figura 25 - Comandos da identificação dos valores dos *pixels*.



Fonte: Aatoria Própria.

Figura 26 - Coleta da amostra classificada.



Fonte: Aatoria Própria.

Figura 27: Tabela de valores das áreas classificadas.

Identificar Resultados				
	Camada	FID	Atributo	Valor
1	Classes		Banda 1	

Exibir Tabela 3 4 Ajuda

Fonte: Aatoria Própria.

O objetivo consistiu em mapear a área de todas as classes com resultados em pixels. São selecionados ou visualizados um dos itens a seguir de modo que o usuário queira ver sua visualização da área, veja os itens a seguir de visualização de área do Qgis.

- **mi**: área em milhas quadradas;
- **me**: área em metros quadrados;
- **k**: área em quilômetros quadrados;
- **a**: área em acres;
- **h**: área em hectares;
- **c**: número de pixels;
- **p**: percentual de cobertura.

No caso em estudo, as áreas foram calculadas em *pixels* e depois convertidas em porcentagem por uma regra simples de três matemática para ficar melhor o entendimento das mudanças do rio ao longo do tempo, ou o usuário pode seguir o seguinte procedimento:

- Divide-se o valor da área fragmentada pelo valor da área total;
- Multiplica-se o resultado por 100.

Basicamente, a fórmula resume bem essa tarefa:

$$(\text{"AREA_CLASSE"} / \text{"AREA_TOTAL"}) * 100$$

O resultado corresponde ao percentual de cobertura para cada feição presente na camada vetorial (SUTTON, 2012).

5.4 MODELAGEM HÍDRICA NAS SÉRIES TEMPORAIS E SEU COMPORTAMENTO

A intenção de se ter uma aplicação técnica na modelagem preditiva é uma avaliação na metodologia da série temporal em estudo, pois identifica-se uma estruturação e comportamento de padrões existentes em seus históricos de construção, assim um modelo preditivo é capaz de produzir dados com certa proximidade dos padrões originais da série temporal em estudo (ADEBIYI; ADEWUMI; AYO, 2014).

Segundo Cowpertwait e Metcalfe (2009), existem domínios temporais, cujo foco se destina na escolha adequada do prazo da série temporal para a consignação de estruturas progressivas da série, ou seja, o padrão desta série é usar uma metodologia onde uma certa porta temporal do histórico passado e do presente se abram para uma porta futura com determinações de ações modificadas que possam ajudar a entender informações e conhecimento de certos acontecimentos que possam ocorrer durante o estudo da série (ADEBIYI; ADEWUMI; AYO, 2014).

Tendo em vista a importância da série temporal para se entender as ações futuras, entra em jogo a relação dela com a modelagem hídrica, o estudo de séries temporais é uma empreitada bastante complexa, quando aplicado em variados modelos e domínios.

Assim, Ni e Yin (2009) e Dablemont *et al.*, (2003), nos diz que, alguns estudos mostram que determinados tipos de características de séries temporais, como por exemplo as séries financeiras e hidrológicas, possuem modelos de comportamento que se repetem ao passar da série temporal com o tempo. Esta repetição de comportamento e padrões é de extrema importância para a condição de estudo de melhorias e avanços do estudo da exatidão ou acurácia das previsões, pois estes estudos mostram que estas séries geralmente expõem repetições de comportamento.

A intenção de se estudar a metodologia e análise de séries temporais é a de obter comparações entre duas ou mais séries em estudo, abrindo análise de soluções onde a condição de semelhança entre elas podem auxiliar nos processos de decisão e de qual apresenta melhor classificação das séries (ADEBIYI; ADEWUMI; AYO, 2014; NOVA FILHO; PESSOA, 2015).

Os modelos computacionais e que envolvem a inteligência artificial como Árvore de Regressão, Redes Neurais e Máquinas de Vetores por Suporte (SVMs), possuem certo interesse em estudo e de pesquisadores quando se trata de previsão e modelagem regressiva. Tais modelos conseguem criar e modelar a partir de dados de histórico passado, que podem ser desenvolvidos a outros conjuntos de dados de problemática que não se encontravam no histórico do conjunto em estudo (LAHMIRIS, 2011; FACELI *et al.*, 2011; ADEBIYI; ADEWUMI; AYO, 2014).

Importante ressaltar que todos esses estudos foram feitos na busca de se modelar um conjunto de dados de soluções onde o modelo previsto apresenta futuras soluções para o modelo presente.

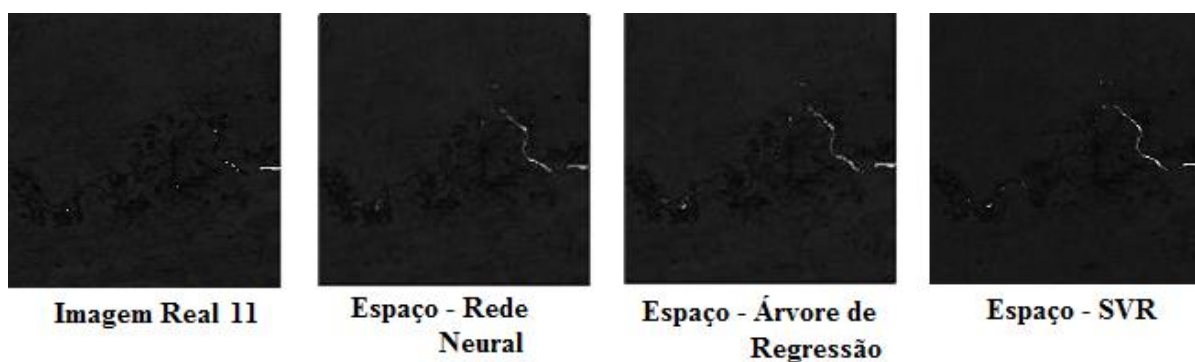
6 ANÁLISES E RESULTADOS

A abordagem de conceitos sob uma nova interpretação, baseada em aprendizado de máquina, para a previsão de imagens de sensoriamento remoto em NDVI e imagens sintéticas, tem por finalidade apresentar modelos que representam a forma semelhante e comportamental do real, bem como apresentar um estudo classificatório e comportamental das mudanças ocorridas no rio Arkansas ao longo do tempo.

6.1 RESULTADOS DAS IMAGENS DE SENSORIAMENTO REMOTO

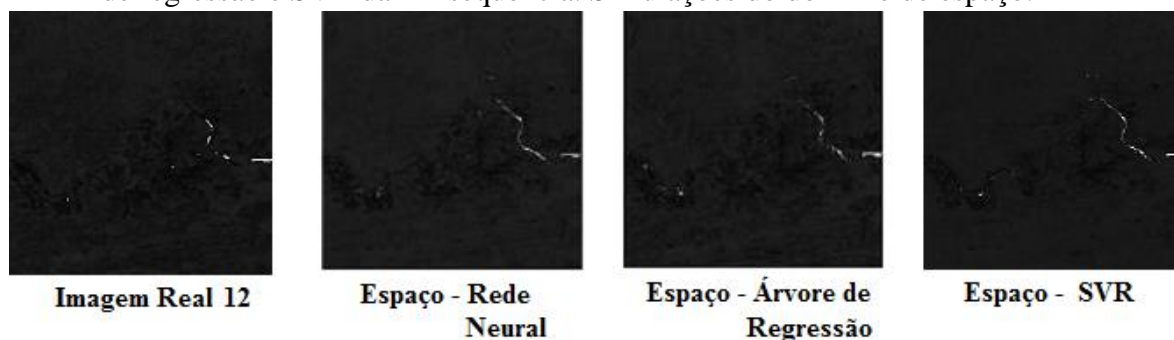
Os resultados apresentados a seguir foram feitos com estratégias e modelagem preditiva, com previsões aplicadas às sequências 11, 12, 13, 14 e 15 das imagens de sensoriamento remoto em NDVI. As imagens foram divididas em imagem real (ideal) e imagens de previsão geradas pelo os algoritmos rede neural, árvore de regressão, SVR. As imagens mostradas pelas figuras de 28 a 32 são de origem bruta sem nenhum tratamento de edição de imagens.

Figura 28 - Imagens resultantes da predição de *pixel* feita pela rede neural e pela árvore de regressão e SVR da 11^a sequência. Simulações do domínio de espaço.



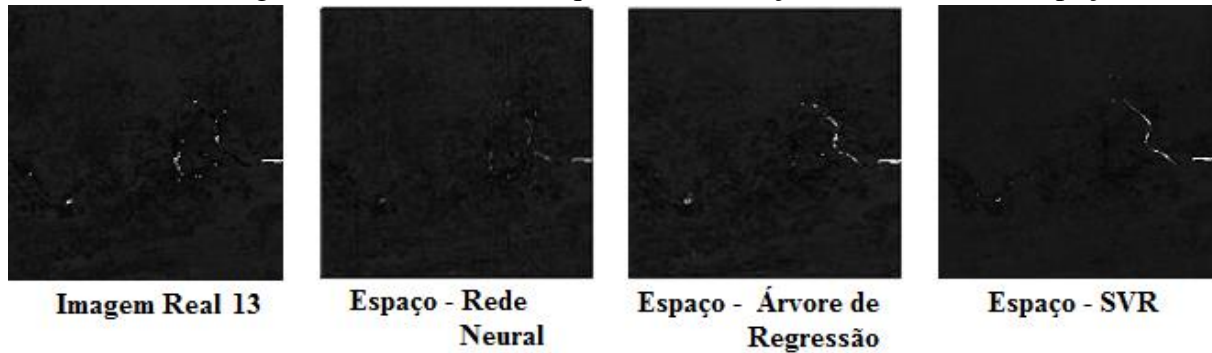
Fonte: Autoria Própria.

Figura 29- Imagens resultantes da predição de *pixel* feita pela rede neural e pela árvore de regressão e SVR da 12^a sequência. Simulações do domínio de espaço.



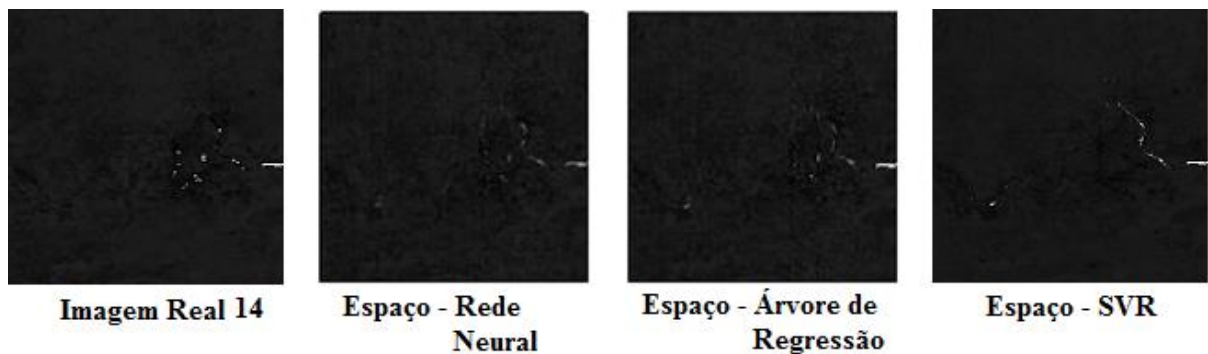
Fonte: Autoria Própria.

Figura 30 - Imagens resultantes da predição de *pixel* feita pela rede neural e pela árvore de regressão e SVR da 13ª sequência. Simulações do domínio de espaço.



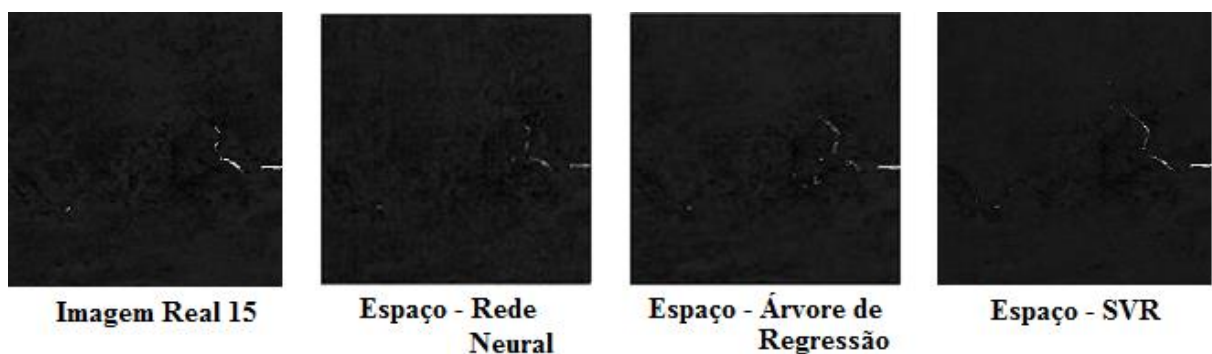
Fonte: Autoria Própria.

Figura 31 - Imagens resultantes da predição de *pixel* feita pela rede neural e pela árvore de regressão e SVR da 14ª sequência. Simulações do domínio de espaço.



Fonte: Autoria Própria.

Figura 32- Imagens resultantes da predição de *pixel* feita pela rede neural e pela árvore de regressão e SVR da 15ª sequência. Simulações do domínio de espaço.



Fonte: Autoria Própria.

As imagens acima mostram que os resultados obtidos com os algoritmos e o uso das imagens em NDVI é vantajoso, uma vez que essas imagens, por ser proporcional à biomassa existente, são mais indicadas quando pretende-se fazer comparações ao longo do tempo em uma determinada área de estudo. O resultado dessa razão também é uma imagem monocromática em escala de cinza, com valores entre 1 e -1 (MENEZES; ALMEIDA, 2012). Quanto mais

próximo de 1, indicará a presença de alta concentração de biomassa; quanto mais próximo de -1 mostrará a ausência de biomassa.

O uso de interpretação quantitativa das imagens em NDVI é relevante para estes levantamentos, pois, com o uso de algoritmos que provam uma nova abordagem baseada em aprendizado de máquina e com intuito de representar de forma semelhante o comportamento real e previsto de ações futuras. Os resultados obtidos neste trabalho demonstram que tal abordagem permite alcançar uma representação satisfatória da disposição da imagem real e estimular novos estudos a fim de comparar o custo computacional da abordagem proposta com outras existentes.

Com isso, a razão para usarmos tais medidas é a necessidade em mensurar a qualidade dos modelos e também o poder de previsibilidade do modelo. A escolha da métrica é determinada na prática, dependendo de qual seja o objetivo, ou seja, como deseja usar este modelo e suas restrições técnicas, matematicamente ou computacionalmente. No nosso caso, seriam as previsões de imagens de sensoriamento remoto em NDVI e as métricas utilizadas foram MAE e RMSE. O MAE (valor Absoluto Médio) é um valor absoluto e menos sensível a valores extremos, em comparação com a função quadrada. Minimizar o MAE é encontrar uma aproximação para o valor mediano do resultado condicional da próxima observação (prevista) dadas as variáveis explicativas (o passado), em vez da expectativa condicional. Já o uso do RMSE (Raiz do Valor do Quadrático Médio), minimizando essa métrica, tenta encontrar uma aproximação para o valor esperado condicional da próxima observação (a ser prevista), dadas as variáveis explicativas (o passado na série temporal).

O MAE e o RMSE podem ser usados juntos para diagnosticar a variação nos erros em um conjunto de previsões. Tanto o MAE quanto o RMSE podem variar de 0 a ∞ . Uma vez que eles são pontuados com orientação negativa, valores menores são melhores (Lima, 2009).

A seguir, mostraremos os resultados obtidos com uso dos três algoritmos (Rede Neural, Árvore de Regressão, SVR) com as métricas MAE e RMSE. As tabelas II e III a seguir mostram o desempenho da Rede Neural, Árvore de Regressão e o SVR com estratégias de cálculo, feitas por meio do cálculo do erro MAE e RMSE.

Na Tabela II, tanto a árvore de regressão como o SVR, na métrica MAE, apresentaram uma acurácia melhor de maneira individual para cada teste, quando o treinamento e regressão das sequências NDVI foi estimada no domínio espacial *pixel-pixel* (x, y) em relação à rede neural. O MAE indicou o afastamento (desvio) médio absoluto dos valores previstos em relação aos valores observados das imagens numa varredura de pixels. Em síntese, existe uma

pontuação linear satisfatória dos resultados de previsões do modelo real para o previsto. O ideal seria que os valores do MAE, fossem o mais próximo de zero.

Tabela II: ERRO MAE (%) DAS PREDIÇÕES (*Pixel vs Pixel*)

Predições no Espaço feitas com a Rede Neural, Árvore de Regressão e o SVR			
Imagem	MAE (%) RN do Espaço	MAE (%) AR do Espaço	MAE (%) SVR do Espaço
Imagem Prevista 11	3,89	1,39	1,39
Imagem Prevista 12	4,27	1,77	1,09
Imagem Prevista 13	4,29	1,61	1,33
Imagem Prevista 14	4,23	1,31	1,25
Imagem Prevista 15	4,10	1,26	1,02

Fonte: Autoria Própria

Tabela III: RMSE (%) DAS PREDIÇÕES (*Pixel vs Pixel*).

Predições no Espaço feitas com a Rede Neural, Árvore de Regressão e o SVR			
Imagem	RMSE (%) RN do Espaço	RMSE (%) AR do Espaço	RMSE (%) SVR do Espaço
Imagem Prevista 11	1,739	0,621	0,621
Imagem Prevista 12	1,909	0,791	0,487
Imagem Prevista 13	1,918	0,720	0,594
Imagem Prevista 14	1,891	0,585	0,559
Imagem Prevista 15	1,833	0,563	0,456

Fonte: Autoria Própria.

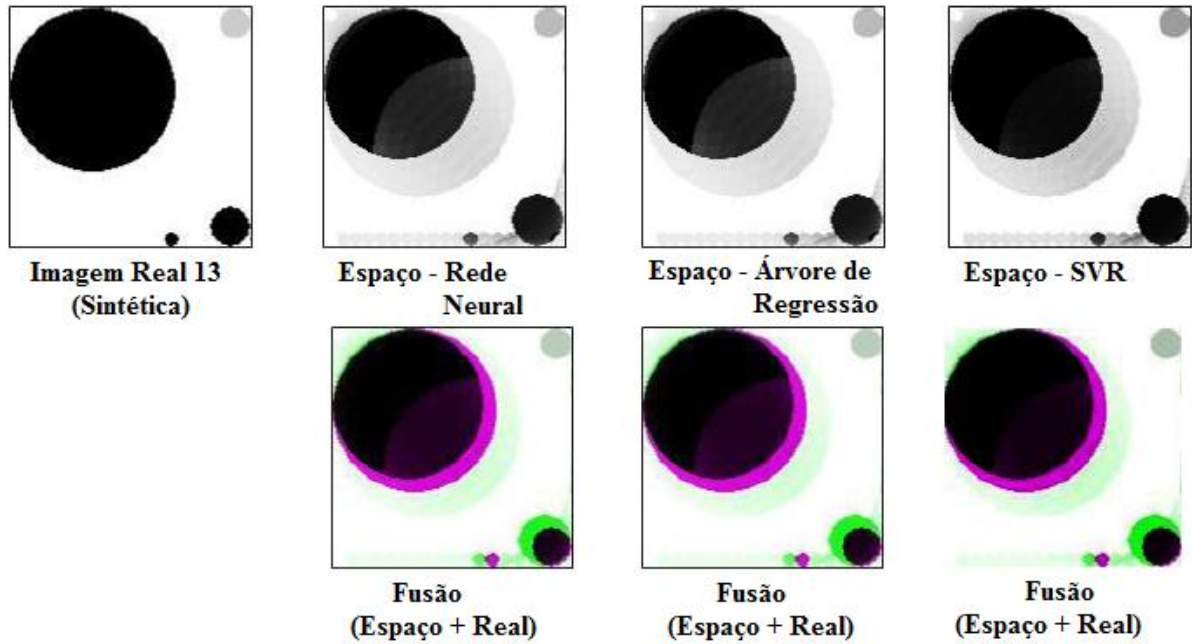
Na Tabela III, os resultados do RMSE foram feitos de maneira a se obter a média entre os resultados dos testes para cada algoritmo, visto que essa métrica, por conter pesos (quantidade de teste), é relativamente boa em relação a grandes erros. Isso significa que o RMSE é útil quando se deseja ter uma precisão do modelo de previsão, evitando assim erros grandes, particularmente indesejáveis. Também como o MAE, o ideal seria que os valores dos resultados fossem próximo de zero, porém, não descartando a eficácia para cada algoritmo utilizado neste trabalho. As métricas utilizadas neste trabalho mostraram a eficácia do desempenho dos algoritmos em fazer previsões, no qual cada medida tem vantagens e desvantagens: a RMSE é melhor para mostrar desvios com atribuição da média dos testes; a MAE tem interpretação mais fácil ou individual para cada teste para se obter a mediana dos resultados.

6.2 RESULTADOS DAS IMAGENS SINTÉTICAS

Para validarmos o uso dos algoritmos em estudo, utilizamos uma segunda abordagem, por meio das imagens sintéticas, mostrando as modificações de predição. Nas previsões feitas com a rede neural, na árvore de regressão as modificações foram suaves; porém, as

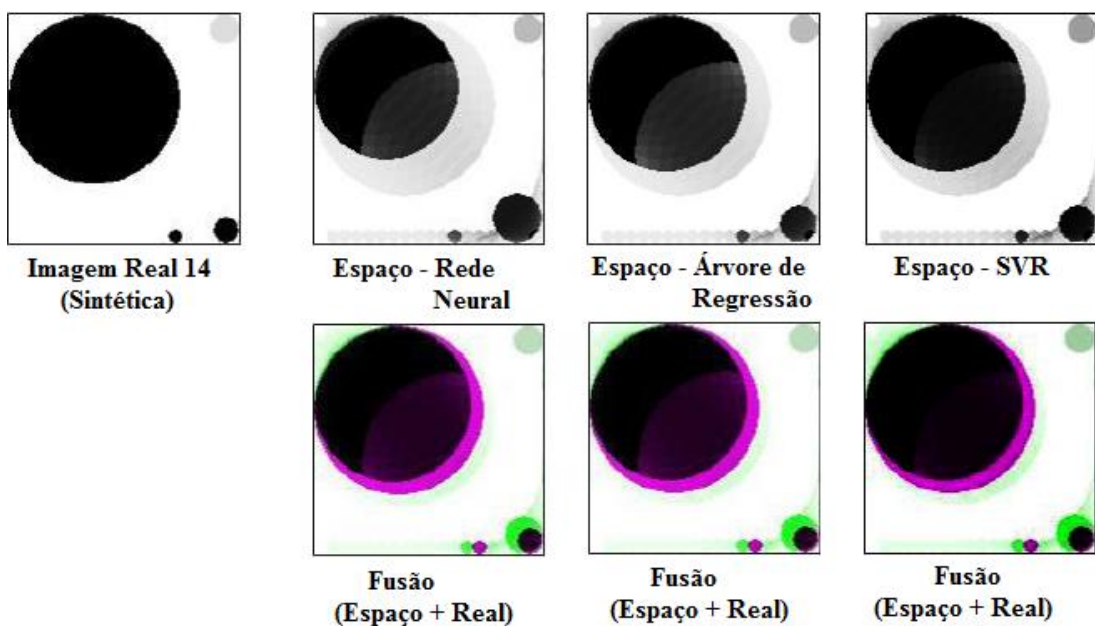
modificações com o SVR ficaram mais visíveis como aumento/diminuição e deslocamentos locais, conforme mostram as figuras de 33, 34 e 35.

Figura 33 - Imagens resultantes da predição de pixel feita pela rede neural da 13ª sequência. As simulações do domínio de espaço A fusão da imagem real com as imagens preditas destaca as diferenças nas cores verde e magenta.



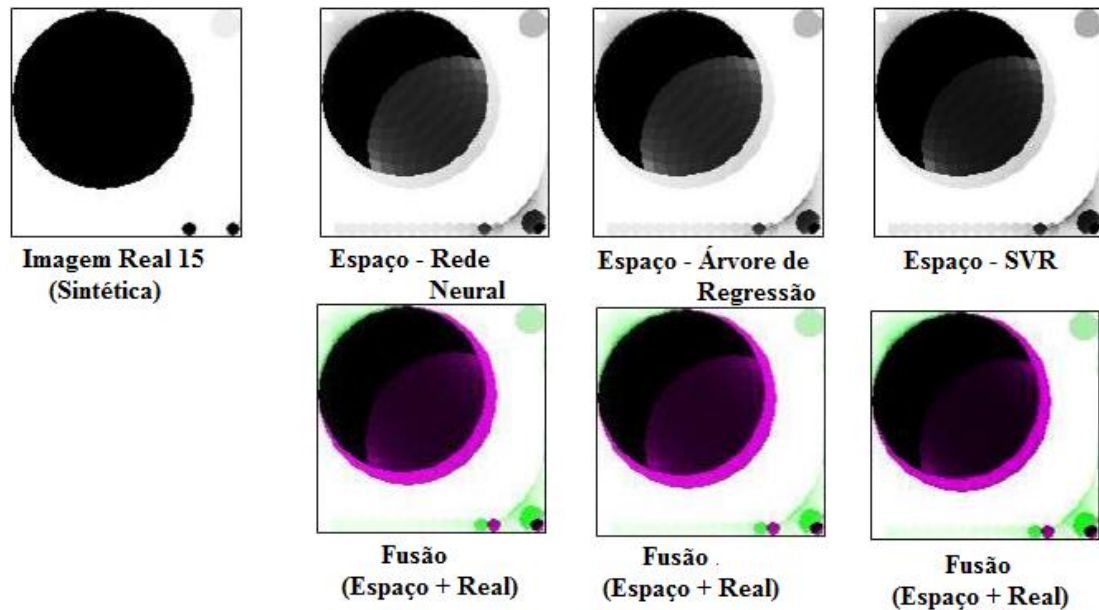
Fonte: Autoria Própria.

Figura 34 - Imagens resultantes da predição de pixel feita pela rede neural da 14ª sequência. As simulações do domínio de espaço A fusão da imagem real com as imagens preditas destaca as diferenças nas cores verde e magenta.



Fonte: Autoria Própria.

Figura 35 - Imagens resultantes da predição de pixel feita pela rede neural da 15ª sequência. As simulações do domínio de espaço A fusão da imagem real com as imagens preditas destaca as diferenças nas cores verde e magenta.



Fonte: Autoria Própria.

As tabelas IV e V a seguir avaliam as predições, por meio dos erros MAE e RMSE, dados em porcentagem, das previsões das imagens 13, 14 e 15. Cada previsão foi feita no domínio do espaço.

Tabela IV: Erro MAE% da predição de *pixel* das imagens 13, 14 e 15 com a Rede Neural, Árvore de Regressão e o SVR.

Predição do Pixel			
Imagem	MAE (%) Rede Neural	MAE (%) Árvore de Regressão	MAE (%) SVR
Imagem Prevista 13	14,28	12,06	11,32
Imagem Prevista 14	12,15	11,63	10,74
Imagem Prevista 15	12,44	10,96	9,23

Fonte: Autoria Própria

Tabela V: Erro RMSE% da predição de *pixel* das imagens 13, 14 e 15 com a Rede Neural, Árvore de Regressão e o SVR.

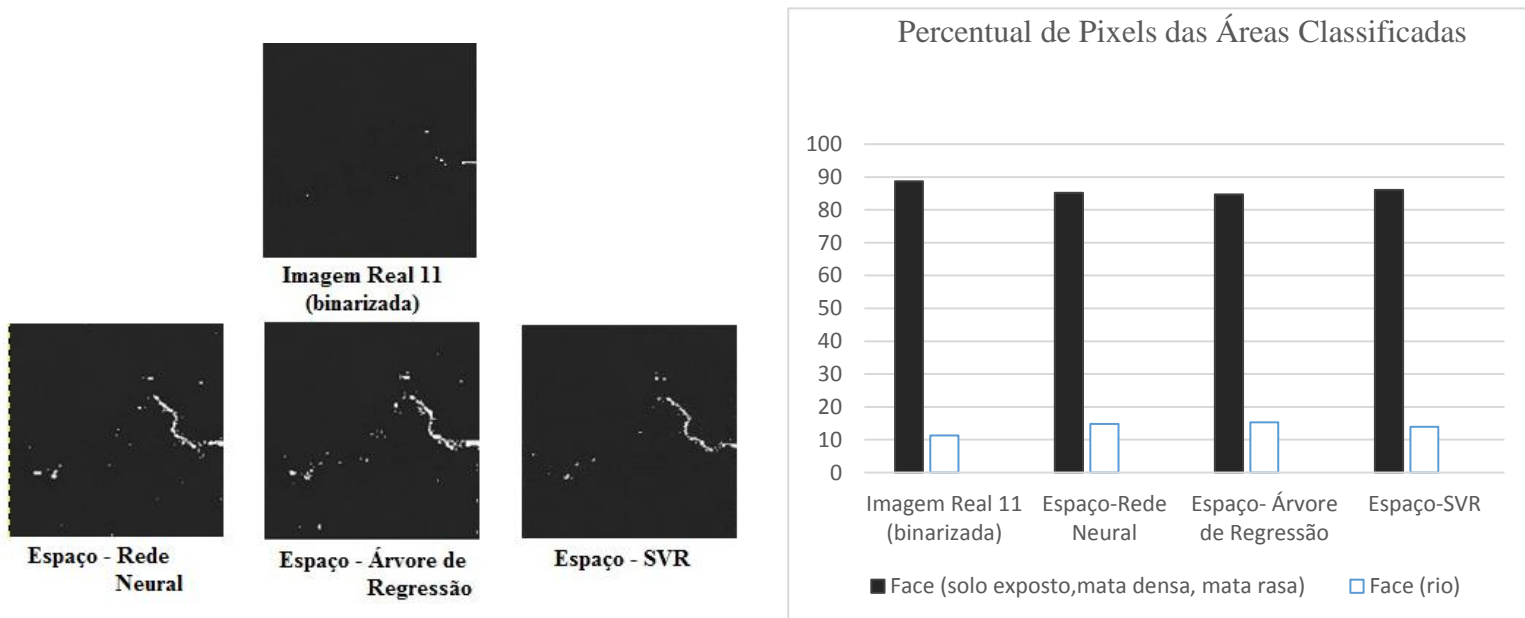
Predição do Pixel			
Imagem	RMSE (%) Rede Neural	RMSE (%) Árvore de Regressão	RMSE (%) SVR
Imagem Prevista 13	8,244	6,961	6,534
Imagem Prevista 14	7,014	6,713	6,199
Imagem Prevista 15	7,180	6,327	5,328

Fonte: Autoria Própria.

6.3 RESULTADOS LOCAIS DAS IMAGENS DE SENSORIAMENTO REMOTO

Outra maneira de avaliar a diferença entre a imagem real e a previsão do rio Arkansas foi pelo cálculo de porcentagem da área dos *pixels* das imagens. Foi utilizada uma binarização, focada somente nas mudanças do rio. Assim, os *pixels* foram classificados em duas faixas de valores: cinza (vegetação densa, vegetação rasa, solo exposto); e, branco (rio), conforme as figuras 36, 37, 38, 38, 40. A classificação foi feita utilizando o *software* Qgis para mapear as áreas dos *pixels* com a mesma cor como nos mostra as tabelas VI, VII, VIII, IX, e X, depois foi feito um gráfico com porcentagem dos níveis das faixas predominante dos *pixels*.

Figura 36 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 11ª imagem.



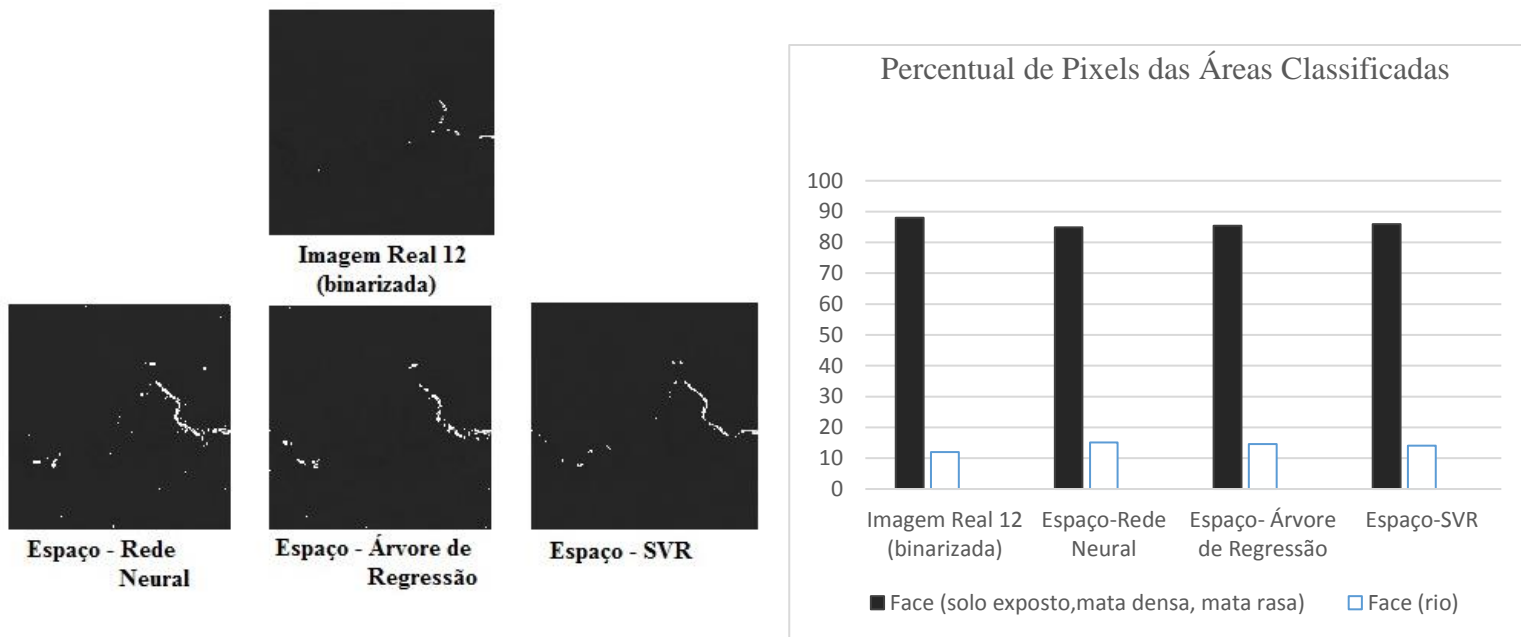
Fonte: Autoria Própria.

Tabela VI: Porcentagem de Pixels nos mapas temáticos associados aos algoritmos e a 11ª imagem.

Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 11(binarizada)	88,74%	11,26%
Espaço- Rede Neural	85,17%	14,83%
Espaço- Árvore de Regressão	84,68%	15,32%
Espaço- SVR	86,11%	13,89%

Fonte: Autoria Própria.

Figura 37 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 12^a imagem.



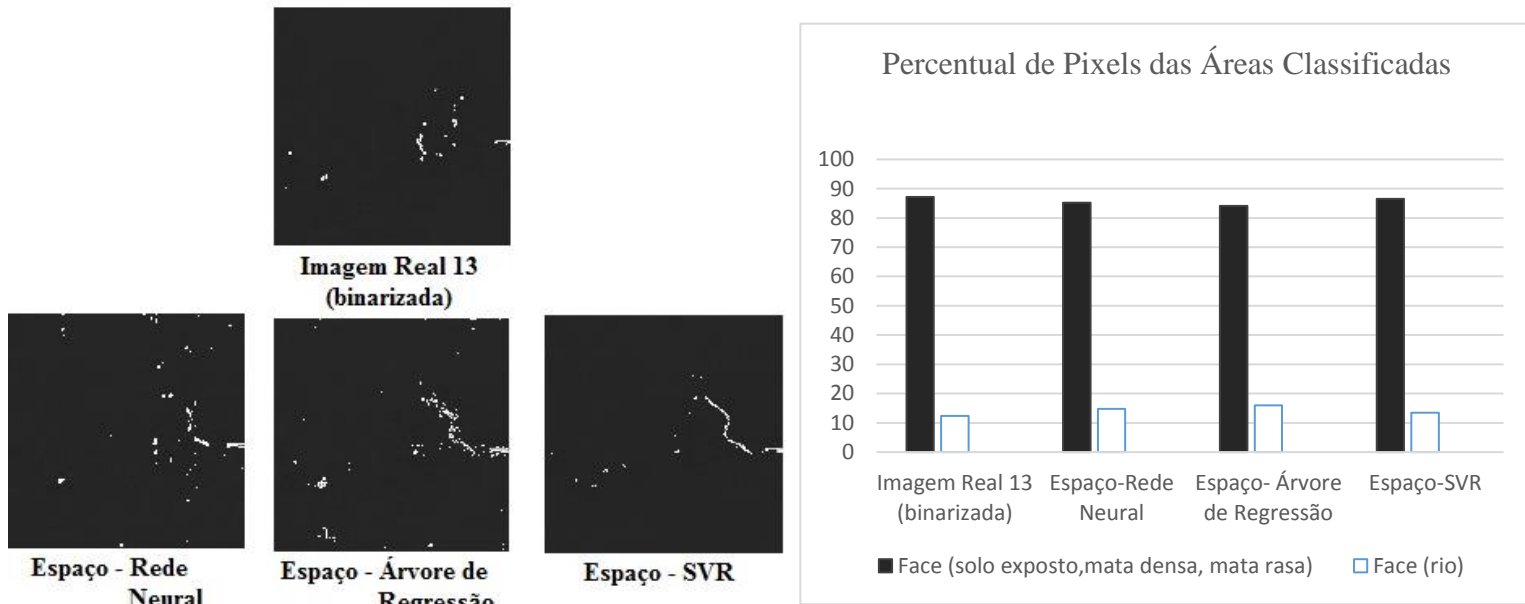
Fonte: Autoria Própria.

Tabela VII: Porcentagem de *Pixels* nos mapas temáticos associados aos algoritmos e a 12^a imagem.

Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 12 (binarizada)	88,02%	11,98%
Espaço- Rede Neural	84,92%	15,08%
Espaço- Árvore de Regressão	85,46%	14,54%
Espaço- SVR	85,98%	14,02%

Fonte: Autoria Própria.

Figura 38 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 13ª imagem.



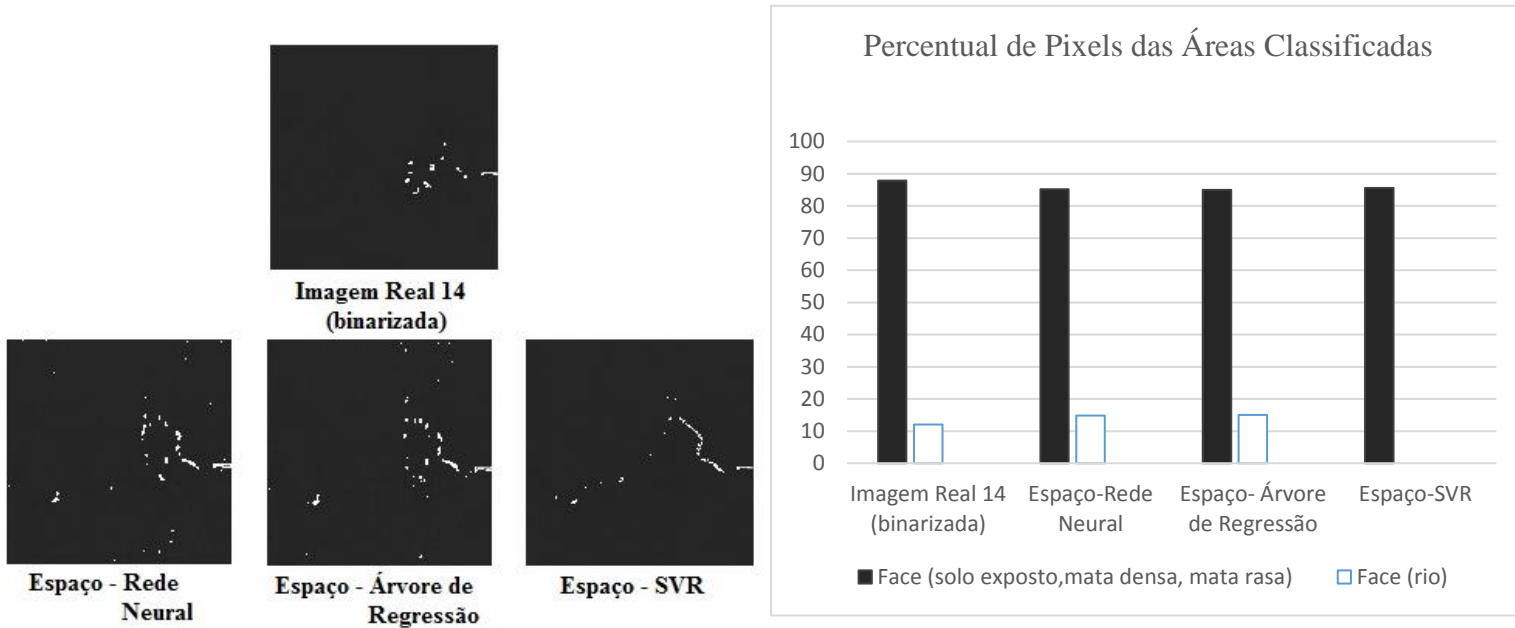
Fonte: Autoria Própria.

Tabela VIII: Porcentagem de *Pixels* nos mapas temáticos associados aos algoritmos e a 13ª imagem.

Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 13 (binarizada)	87,72%	12,38%
Espaço- Rede Neural	85,24%	14,76%
Espaço- Árvore de Regressão	84,08%	15,92%
Espaço- SVR	86,25%	13,75%

Fonte: Autoria Própria.

Figura 39 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 14ª imagem.



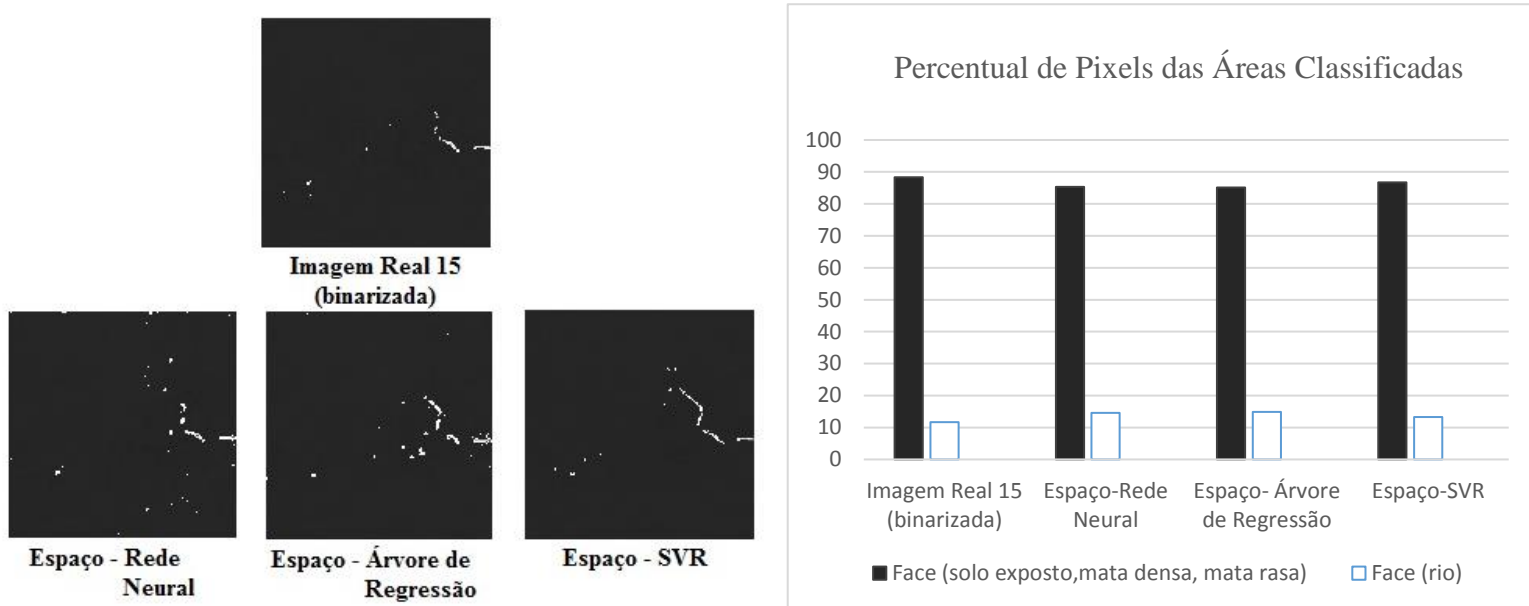
Fonte: Autoria Própria.

Tabela IX: Porcentagem de *Pixels* nos mapas temáticos associados aos algoritmos e a 14ª imagem.

Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 14 (binarizada)	87,91%	12,09%
Espaço- Rede Neural	85,19%	14,81%
Espaço- Árvore de Regressão	84,93%	15,07%
Espaço- SVR	85,66%	14,34%

Fonte: Autoria Própria.

Figura 40 - Mapas Temáticos e níveis de mudanças do rio Arkansas dos algoritmos da 15ª imagem.



Fonte: Autoria Própria.

Tabela X: Porcentagem de *Pixels* nos mapas temáticos associados aos algoritmos e a 15ª imagem.

Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 15 (binarizada)	88,33%	11,67%
Espaço- Rede Neural	85,38%	14,62%
Espaço- Árvore de Regressão	85,13%	14,87%
Espaço- SVR	86,58%	13,42%

Fonte: Autoria Própria.

Os testes foram executados com o mesmo tamanho da área escolhida e diferentes algoritmos de aprendizagem supervisionada, a fim de se fazer uma análise do comportamento do modelo e sua melhoria de previsão de acordo com as mudanças ocorridas. Assim, com a segmentação binarizada e focando nas mudanças do rio Arkansas é possível modelar uma previsão significativa com o algoritmo SVR, pois o mesmo mostrou pelo cálculo de

porcentagem local da área do rio uma previsibilidade de modelos futuros com menores diferenças entre modelo real e o previsto para análise dos impactos de ocorrência do rio.

É relevante observar que o meio hídrico passa por constantes mudanças diárias e ao longo do tempo, isso reforça um estudo em seu comportamento na superfície terrestre, pois, como pertence a uma biomassa onde esse recurso natural é dependente dos fatores locais e climáticos, suas mudanças podem favorecer ou interferir nas vidas existente naquele local.

7 CONSIDERAÇÕES FINAIS

A proporção de comparar o desempenho de três métodos de regressão preditivos, baseados em técnicas de aprendizado de máquina, para previsão de séries temporais de sequências a partir de imagens MODIS processadas em NDVI torna-se uma ferramenta que despertar o interesse na pesquisa devido aos acontecimentos que envolvem a ação do meio ambiente e do ser humano. Este fato permitiu acelerar a eficiência computacional dos algoritmos, pois o processamento dos dados de entrada foi feito diretamente em um espaço as quais foram aplicadas de forma escalonada sobre a série temporal das imagens NDVI originais.

Ao implementar métodos de regressão preditivos baseados em aprendizado de máquina para estimar modelos de previsão de imagens NDVI, visamos construir uma ferramenta computacional inteligente que possa orientar as autoridades públicas da área de gestão ambiental na tomada de decisões, no tocante à prevenção e mitigação de eventuais danos e desastres ambientais, que possam ser produzidos ao ecossistema pelo impacto da complexa interação homem-natureza.

A partir de uma análise do comportamento desses resultados é cabível um estudo em mais profundidade e de interesse das imagens de sensoriamento remoto.

As métricas de erro MAE e RMSE apresentaram resultados globais acurados, quando avaliadas sobre as imagens de saída estimadas dos algoritmos, no espaço de treinamento pixel-pixel neste trabalho. Os resultados obtidos com SVR mostraram-se promissores para estimar modelos de previsão de séries temporais de imagens NDVI de sistemas complexos envolvendo interação homem-natureza.

Percebe-se a importância de se procurar ferramentas ou metodologia computacional com certo grau de confiabilidade para não cair no acaso do uso de métodos que não tenham respaldo científico, assim obtendo resultados seguros que possam ser usados para diversas finalidades científicas.

Esta produção abre novas perspectivas para a implementação de um comitê de máquinas maior, podendo envolver árvore de regressão, rede neural, suporte por vetores de máquina, dentre outros. O uso de um comitê de aprendizado de máquina mais amplo seria motivado pela busca de se construir sistemas inteligentes que permitam a estimação e modelagem acurada do comportamento dinâmico dos sistemas complexos que objetivaram este estudo.

Várias pesquisas foram desenvolvidas na busca de técnicas de aprendizado de máquina na estimativa de previsões e tais técnicas apresentaram resultados muito próximos dos valores

reais, o que permite que projetos de predições futuros tenham maiores chances de sucesso. Ademais, sugere-se que, para trabalhos posteriores, sejam aplicadas essas técnicas em bases de dados com projetos de sistemas web e de dispositivos móveis, com o intuito de comparar os resultados de acurácia.

Estamos envolvidos direta, indireta e fisicamente em mudanças, sendo isso resultado das ocorrências que foram estabelecidas com o uso e preservação dos recursos naturais, até um certo tempo não havia a preocupação com estudos envolvendo o meio hídrico. Hoje, começamos a dar alguns passos para ampliar a consciência e procurar mudar certos comportamentos em relação ao que está ocorrendo para a crescente e alarmante crise ambiental que vivemos, com um crescimento de ações graves que vêm perturbando a população local e mundial.

Num futuro próximo, este trabalho e o aperfeiçoamento para previsão de imagens remotas alicerçará a realização de estudos de caso a partir de dados nativos obtidos no nordeste brasileiro, região que sofre com a escassez de água e que padece com processos ambientais e/ou climáticos agressivos, como a desertificação e a seca. Por conseguinte, a execução satisfatória em solo brasileiro contribuirá para o monitoramento estratégico de seus reservatórios.

7.1 APORTE CIENTÍFICO DA DISSERTAÇÃO

Com o uso dos três algoritmos de regressão com aprendizado de máquina, abre-se uma nova observação e aprimoramento dos métodos para o meio científico como uma inovação no aperfeiçoamento das técnicas utilizadas, assim pondera-se a contribuição da dissertação da seguinte forma:

- A utilização e possibilidade de se usar metodologias para se fazer previsões com imagens hídricas, utilizando series temporais se mostrou viável para o estudo da progressão de mudanças locais como rios, açudes, lagoas que deseje-se estudar para estimar sua falta ou mudanças geográficas;
- O uso de imagens em NDVI mesmo sendo para estudo de vegetação mostrou-se promissor quando utilizado para o estudo de caso de mudanças hídricas;

- Classificar imagens envolvendo recurso hídrico de certa forma é uma pesquisa complexa, pois é um recurso que constantemente passa por mudanças e para se modelar uma previsão dela requer uma investigação e pesquisa a fundo.

7.2 TRABALHOS FUTUROS

Como desenvolvido no estudo de caso sobre pesquisa hidrológicas futuramente pretende-se fazer trabalhos e aperfeiçoamento para previsão de imagens remotas alicerçará a realização de estudo de caso a partir de dados nativos obtidos no nordeste brasileiro, região que sofre com a escassez de água e que padece com processos ambientais/climáticos agressivos, como a desertificação. A execução satisfatória em solo brasileiro contribuirá para o monitoramento estratégico de seus reservatórios.

Fazer a implementação de um comitê de máquinas com o SVR, árvore de regressão e a rede neural. Montar um comitê de máquinas, consiste em fazer a junção de vários aprendizados de máquinas na solução de uma única causa de estudo em meios computacionais em outros casos de estudo. Com isso será utilizado este comitê com a mesma metodologia, utilizando séries temporais para a acurácia dos resultados.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMCIK, F.; BRÉDA, R.; VAISPACHER, T. **Analysis of GPS satellites position determination using MATLAB® environment**. IEEE. Military Technologies (ICMT), 2015. International Conference on, 2015.
- ADEBIYI, A. A.; ADEWUMI, A. O.; AYO, C. K. **Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction**. Journal of Applied Mathematics, v.2014, 2014.
- AKBAR, H. et al. **Training neural networks using Clonal Selection Algorithm and Particle Swarm Optimization: A comparisons for 3D object recognition**. IEEE. Hybrid Intelligent Systems (HIS), 2011 11th International Conference on, 2012.
- AMORIM NETO, M. et al. **Improving financial time series prediction using exogenous series and neural networks committees**. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, Barcelona, Spain. Proceedings. IEEE, 2010.
- ARAÚJO, S. A. **Métodos inteligentes para previsão de séries temporais em imagens de sensoriamento remoto NDVI**. Dissertação (Mestrado em Mestrado em ciências da computação UERN-UFERSA). Mossoró: PPgCC UERN-UFERSA, 2016.
- BARBOSA, J. P. M. **UPBox: Armazenamento na Nuvem para Dados de Investigação da U. Porto**. Tese (Doutorado) — Universidade do Porto, 2013.
- BARKER, J.L. *et al.* **MODIS Spectral Sensivity Study: Requirements and Characterization**. p- 84.Out. 1992.
- BISHOP, C. M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006.
- BOSE, B. E.; GUYON, I. M.; VAPNIK, V. N. **A Training Algorithm for Optimal Margin Classifiers**. In: ANNUAL WORKSHOP ON COMPUTATIONAL LEARNING THEORY, New York, NY, USA. Proceedings. ACM Press, 1992.
- CAMARA, G.; DAVIS, C.; MONTEIRO, A. **Introdução a Ciência da Geoinformação**. [S.l.: s.n.], 2001.
- CARVALHO C. M.; ALMEIDA FILHO, R. **Assessment of desertification in the southwestern state of piauí, with remote sensing techniques**. Brazilian Symposium on Remote Sensing, 2007.
- CHRISTOFIDIS, D. **A água e a crise alimentar**. 1997. Disponível em <www.iica.org.br/Aguatrab/Demetrios%20Christofidis/P2TB01.htm>. Acesso em 03 de setem. de 2017.
- CORTES, C. VAPNIK, V. 1995. **Support Vector Networks**, Machine Learning. p: 273–297.nov. 1995.
- COWPERTWAIT, P. S. P.; METCALFE, A. V. **Introductory Time Series with R**. 1st.ed. New York, NY, USA: Springer-Verlag, 2009.

- DABLEMONT, S. *et al.* **Time series forecasting with SOM and local non-linear models - Application to the DAX30 index prediction.** In: WORKSHOP ON SELF-ORGANIZING MAPS, Kitakyushu, Japan. Proceedings. 2003.
- DANTAS, D. N.; *et al.* **Previsão de série temporal de sinais sísmicos avo utilizando redes neurais.** Simpósio Brasileiro de Automação Inteligente, 2013.
- DE MYTTENAERE, B. *et al.* **“Mean absolute percentage error for regression models”.** Neurocomputing, vol. 192, 2016.
- FACELI, K. *et al.* **Inteligência artificial: Uma abordagem de aprendizado de máquina.** Rio de Janeiro: LTC, v. 2, p. 192, 2011.
- FENTIS, A. *et al.* **Previsão de Energia Solar de Curto Prazo usando Regressão por Vetores de Suporte e Feed-Forward NN.** New Circuits and Systems Conference (NEWCAS), 2017 15th IEEE International, 2017.
- NOVA FILHO, V.; PESSOA, S. R. **Previsão de séries temporais utilizando pools de preditores criados a partir do particionamento da série e da divisão da tarefa de previsão.** Dissertação (Mestrado em Mestrado em ciências da computação - UFPE). Recife, 2015.
- GCT-PH1 – Population, Housing Units, Area, and Density: 2010 – state – Place and (in selected states)country.2011.Disponívelem :https://geonames.usgs.gov/apex/f?p=gnispq:3:0::NO::P3_FID:78956. Acesso em 20 de mai. 2018.
- GILAT, AMOS. **Matlab com Aplicações em Engenharia.** Bookman Editora Ltda; empresa do grupo Educação S.A. Porto Alegre – RS, 2012.
- GONÇALVES, A. R. **Máquina de vetores suporte.** Unicamp. 2015. Disponível em: <https://www.researchgate.net/profile/Andre_Goncalves2/publication/266411199_Maquina_de_Vetores_Suporte/links/54b664b00cf2bd04be3206bd/Maquina-de-Vetores-Suporte.pdf >. Acesso em 20 de mai. 2018.
- HAYKIN, S. **Neural networks: A comprehensive foundation.** Prentice Hall International, 2001.
- LAHMIRI, S. **A comparative study of backpropagation algorithms in financial prediction.** International Journal of Computer Science, Engineering and Applications (IJCSA), Montreal, QC, Canada, v.1, n.4, p.15–21, Aug. 2011.
- LATORRE, M, L. *et al.* **Sensor Modis: características gerais e aplicações. Espaço & Geografia,**2003. < <http://www.lsie.unb.br/espacoegografia/index.php/espacoegografia/article/view/111/109> > Acesso em 13 de fevereiro de 2017.
- LIMA, C. M. M. **Um estudo sobre o método Mínimos Quadrados Móveis por Aproximações Iteradas.** Pontifícia Universidade Católica do Rio de Janeiro,(PUC-RJ). Rio de Janeiro. 2009.
- HASTIE, T.; *et al.* **An introduction to statistical learning.** Springer, 2013.

KANAN, C.; COTTRELL, G. W. **Color-to-grayscale: Does the method matter in image recognition**. Department of Computer Science and Engineering, University of California San Diego. Califórnia, United States of America, 2012.

KWOK, J. T.-Y. **Support vector mixture for classification and regression problems**. In: IEEE. Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on. 1998.

KUMAR, D. A.; ALI, H. A.; HAMMED, S. **Detecting diseased images by segmentation and classification based on semi-supervised learning**. 12th International Conference on Hybrid Intelligent Systems (HIS), 2012.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento digital de imagens**. Rio de Janeiro: Brasport, 1999.

MARTINEZ, A. P. **Uso eficiente del agua en riego**. In: **Uso eficiente del agua**. 1991 <www.unesco.org/uy/phi/libros/uso_eficiente/cap4.html> Acesso em 13 de fevereiro de 2018.

MEDEIROS, Anderson C. L., et al. **Produção de Mapas e Geoprocessamento**. Workshop Brasileiro de Produção de Mapas, 2018.

MENEZES, P.; ALMEIDA, T. **Introdução ao processamento de imagens de sensoriamento remoto**. UNB, EBook, Brasília, 2012. Disponível em: <<http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9f-b82a-e9693e4f69d8>>. Acesso em 28 mai. 2018.

MERINO, R, F; NUNEZ, M; **Self-Adaptive Induction of Regression Trees**. IEEE. **Transactions on Pattern Analysis and Machine Intelligence**. (Volume: 33, Issue: 8, Aug. 2011). 2011.

MISAGHI, S. *et al.* **A hybrid model based on support vector regression and modified harmony search algorithm in time series prediction**. IEEE. Fuzzy and Intelligent Systems (CFIS), 2017 5th Iranian Joint Congress on, 2017.

MCCLELLAN, KERR. **Arkansas River Navigation System (MKARNS)**". History & Culture. The Encyclopedia of Arkansas, 2010.

MODIS NDVI data for Texas from CSR em: <<http://web.corral.tacc.utexas.edu/CSR/Public/NDVI-MODIS>>. Acesso em 20 de novembro de 2016.

MORAES, E.C.; Capítulo 1: **Fundamentos de Sensoriamento Remoto**. São José dos Campos: INPE. 2002.

MOREIRA, M. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. [S.l.]: Universidade Federal de Viçosa, 2005.

NASA. MODIS web. 2007 <[Http://tpwww.gsfc.nasa.gov/MODIS/index.html](http://tpwww.gsfc.nasa.gov/MODIS/index.html)> Acesso em 20 dezembro 2016.

NI, H.; YIN, H. **Exchange rate prediction using hybrid neural networks and trading indicators**. Neurocomputing, [S.l.], v.72, n.1315, p.2815–2823, Aug. 2009.

NORVIG, P.; RUSSELL, S. **Inteligência Artificial**, 3a Edição Elsevier Brasil, 2014.

NOVAKOVIC, J.; VELJOVIC, A. **C-support vector classification: Selection of kernel and parameters in medical diagnosis**. In: Intelligent Systems and Informatics (SISY), 2011. IEEE 9th International Symposium on. p. 465–470. 2011.

PAL, N. R.; PAL, S. K. **A review on image segmentation techniques**. Science Direct. Pattern Recognition 1993.

QUEIROZ, J. E. R. de; GOMES, H. M. **Introdução ao processamento digital de imagens**. Departamento de Sistemas e Computação. Revista RITA, UFCG, 2001.

QUINLAN, J. R. **Combining instance-based and model-based learning**. In: **Proceedings of the Tenth International Conference on Machine Learning**. [S.l.: s.n.], 1993.

RADKE, R.J.; ANDRA, S.; AL-KOFAHI, O.; ROYSAM, B. **Image change detection algorithms: a systematic survey**. IEEE Transactions on Image Processing, v. 14, n. 3, p. 294–307, 2005.

RAJAGOPALAN, S.; *et al.* **Evaluation of thresholding techniques for segmenting scaffold images in tissue engineering**. Proceedings of SPIE - The International Society for Optical Engineering, 2004.

RIBEIRO, G. *et al.* **Lag selection for time series forecasting using Particle Swarm Optimization**. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, San Jose, California, USA. Proceedings. IEEE, 2011.

RODRIGUES, A. R. *et al.* **Prediction System of Bus Arrival Time Based on Historical Data Using Regression Models**. Proceedings of the XI Brazilian Symposium on Information Systems. 2015.

ROUSE, J.W., HAAS, R.H., SCHELL, J.A. and DEERING, D.W. **Monitoring vegetation systems in the Great Plains with ERTS**. In 3rd ERTS Symposium, NASA SP-351 I, pp. 309–317. 1973.

RUSSEL, S.; NORWING, P. **Artificial intelligence a modern approach second edition—upper saddle river**. NS: Prentice Hall/2003, 2003.

SMOLA, A. J.; SCHÖLKOPF, B. **A Tutorial on Support Vector Regression**. Manufactured in The Netherlands, 2004.

SILVA, L.; COSTA FILHO, J. F.; LIMA, E. R. V.; ALMEIDA, H. A. **Analysis of Albedo and Normalized Difference Vegetation Index (NDVI) in a Watershed, using SEBAL - Surface Energy Balance Algorithm for Land**. In: XIII Brazilian Symposium on Applied Physical Geography. Viçosa-MG: Federal University of Viçosa, 2009.

SANTOS, E. M. **Teoria e aplicação de support vector machines à aprendizagem e reconhecimento de objetos baseado na aparência.** Tese (Doutorado) — Universidade Federal da Paraíba, 2002.

SANTOS, C. G. P. **Inversão acústica de dados sísmicos pós empilhamento através de redes neurais recorrentes.** Tese(Doutorado) — Universidade Federal do Rio de Janeiro, 2008.

SENTHILKUMARAN, N.; VAITHEGI, S. **Image segmentation by using thresholding techniques for medical images.** Computer Science & Engineering: An International Journal (CSEIJ), 2016.

SILVA, F. de A. **Aplicação das Técnicas de Processamento Digital de Imagens na Caracterização de Materiais.** Tese (Doutorado) — Unicamp, São Paulo, 1996.

SHUMWAY, R. H.; STOFFER, D. S. **Time Series Analysis and Its Applications - with RE xamples.** 3rd.ed. New York, NY, USA: Springer-Verlag, 2011.

SUTTON, TIM. **Manual de Treinamento do Qgis.** Gentle Introduction to GIS. 2012. Disponível em: <https://docs.qgis.org/2.8/pt_BR/docs/training_manual/foreword/foreword.html>. Acesso em 26 abr. 2018.

TEIXEIRA, L. L. *et al.* **Comparativo entre arima, redes neurais artificiais, máquinas de vetores suporte para regressão e combinação linear de previsões de vazões mensais no posto 266-itaipu.** XXI SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO. Bauru- SP, 2014

VAPNIK, V. N. **The Nature of Statistical Learning Theory.** New York: Springer – Verlag. 1995.

VIDHYA, M. **Efficient classification of portscan attacks using support vector machine.** In: Green High Performance Computing (ICGHPC), 2013 IEEE International Conference on. [S.l.: s.n.], 2013.

YIGIT,M, E. *et al.* **Abordagem Baseada na Genética para a Síntese de uma Antena Conformal Cilíndrica-Retangular de Microcópica usando Redes Neurais Artificiais e Modelos de Regressão por Vetores de Suporte.** IEEE. Electrical and Electronics Engineering (ELECO), 2017 10th International Conference on, 2018.

ZUBEN, F.; ATTUX, R. **Computação Evolutiva. Material Complementar da disciplina IA013–Introdução à Computação Natural.** [S.l.]: Universidade de Campinas– UNICAMP. October, 2008.

APÊNDICE

Códigos

Rede Neural:

```
function [ ] = PreditorPixel(localpasta, formatoArquivo, id,
NumIteracoes, inputDelays, feedbackDelays, numNeuronios,
divisaoTreino, previsao_Espaco, compactar)

mkdir(localpasta, 'Pixel');
localpastaPixel = strcat(localpasta, 'Pixel\');

netInstancia =
narxnet(inputDelays, feedbackDelays, numNeuronios)

view(net);
    netInstancia.trainParam.epochs = 700;%Maximum number of
epochs to train
    netInstancia.trainParam.goal = 0;%PerformaNCSe goal
    netInstancia.trainParam.max_fail = 5;%Maximum validation
failures
    % net.trainParam.mem_reduc = 1;%Factor to use for
memory/speed tradeoff
    netInstancia.trainParam.min_grad = 1e-10;%Minimum
performaNCSe gradient
    netInstancia.trainParam.mu = 0.001;%Initial mu
    netInstancia.trainParam.mu_dec = 0.1;%mu decrease factor
    netInstancia.trainParam.mu_inc = 10;%mu decrease factor
    netInstancia.trainParam.mu_max = 1e10;%Maximum mu
    netInstancia.trainParam.show = 25;%Epochs between displays
(NaN for no displays)
    netInstancia.trainParam.showCommandLine = 0;%Generate
command-line output
    netInstancia.trainParam.showWindow = 0;%Show training GUI
    netInstancia.trainParam.time = inf;%Maximum time to train
in seconds

    netInstancia.divideFcn = divisaoTreino;

'resultados\', num2str(N), 'seq\');

for iteracoes = 0:NumIteracoes

    % idaux evita a repetição de nomes de arquivo, indo de id
até
    % id+Niterações
```

```

    idaux = id + iteracoes;

'C:\Users\Windows\Documents\E\2018\ProjetoMestrado\serieTemporal\serieEspacoTemporal\COMPR-Nivel2.jpg';

matriz de células IM com toda a série temporal das imagens
IM = cell(N); % prealocação de memória

for i=1:N, % para i de 1 até o tamanho da série, armazena as
imagens no vetor IM

    IM{i} = rgb2gray(imread(strcat(localpasta,
sprintf('%d.%s',i, formatoArquivo))));% imagem i da série
temporal

end

fprintf('Lendo as imagens \n');

cria os coeficientes da transformada wavelet para todas as
imagens da
% série temporal

cA = cell(N); % prealocação de memória
cH = cell(N); % prealocação de memória
cV = cell(N); % prealocação de memória
cD = cell(N); % prealocação de memória

for i=1:N, % para i de 1 até o tamanho da série, armazena os
coeficientes nos vetores de células cA, cH, cV, cD

fprintf('Convertendo para grayscale \n');

map = gray(255);

for i=1:N,

    imwrite(uint8(cA{i}), strcat(localpastaPixel,
sprintf('nivel1\cA%d.jpg',i)));
    imwrite(uint8(cH{i}), strcat(localpastaPixel,
sprintf('nivel1\cH%d.jpg',i)));
    imwrite(uint8(cV{i}), strcat(localpastaPixel,
sprintf('\nivel1\cV%d.jpg',i)));
    imwrite(uint8(cD{i}), strcat(localpastaPixel,
sprintf('\nivel1\cD%d.jpg',i)));

end

```

Árvore de Regressão

```
clear
localpasta = 'C:\Users\Rômulo\OneDrive\Imagens\Rio\Nova pasta
(3)\Imagens\';
cd (localpasta);
% id evita que o arquivo seja salvo com o mesmo nome na mesma
pasta
N = 11; %tamanho da série temporal
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
<<<<<<<<
DCT = false;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% matriz de células IM com toda a série temporal das imagens
IM = cell(N); % prealocação de memória

fprintf('Lendo as imagens \n');
for i=1:N, % para i de 1 até o tamanho da série, armazena as
imagens no vetor IM
    %lá em baixo o IDCT!!!!!!
    if (DCT == true)
        IM{i} = dct2(rgb2gray(imread(strcat(localpasta,
sprintf('%d.jpg',i)))));% imagem i da série temporal
    elseif (DCT == false)
        IM{i} = rgb2gray(imread(strcat(localpasta,
sprintf('%d.jpg',i)))));
    end
end
end
fprintf('Imagens Lidas\n');
fprintf('Convertendo para o dominio de espaço \n');

% tamanho das imagens (serve para delimitar o tamanho das
matrizes - retorna [Número de linhas, número de colunas]
fprintf('Calculando tamanho das imagens \n');
[NL,NC]=size(IM{1});
fprintf('Imagens de %i por %i \n',NL, NC);
% cria a imagem de saída
Y = zeros(NL,NC); % inicializa a imagem de saída com
zeros(Número de linhas, número de colunas).

fprintf('Realizando o treinamento \n');
tic
barra = waitbar(0, 'Treinando...');
% Varre todas as imagens, fazendo a previsão de cada pixel,
até montar a imagem prevista
for i=1:NL,

    for j=1:NC,
```

```

%barra
waitbar(i / NL);
%barra

P = double([
    IM{1}(i,j);
    IM{2}(i,j);
    IM{3}(i,j);
    IM{4}(i,j);
    IM{5}(i,j);
    IM{6}(i,j);
    IM{7}(i,j);
    IM{8}(i,j);
    IM{9}(i,j);
    IM{10}(i,j);
    %IM{11}(i,j);
    %IM{12}(i,j);
    %IM{13}(i,j);
    %IM{14}(i,j);
]);

T = double([
    IM{2}(i,j);
    IM{3}(i,j);
    IM{4}(i,j);
    IM{5}(i,j);
    IM{6}(i,j);
    IM{7}(i,j);
    IM{8}(i,j);
    IM{9}(i,j);
    IM{10}(i,j);
    IM{11}(i,j);
    %IM{12}(i,j);
    %IM{13}(i,j);
    %IM{14}(i,j);
    %IM{15}(i,j);
]);

rtree = RegressionTree.fit(P,T);
Y(i,j) = predict(rtree, mean(IM{N}(i,j)));

end
end

close(barra); %fecha a barra

%Tempo de treinamento
timerTreinamento = toc;

```



```
fprintf('O tempo gasto foi de %f minutos\n',
timerTreinamento/1200);

if (DCT == true)
    Y = idct2(Y);
    imwrite(uint8(Y), strcat(localpasta,
sprintf('%dDCTTree.jpg',N)));
elseif(DCT == false)
    imwrite(uint8(Y), strcat(localpasta,
sprintf('%dEspacoTree.jpg',N)));
end
%Y = idct2(Y);
% salva a imagem prevista

fprintf('Processo concluído \n');
```

“SVR” e “erro MAE”

```

num_image = 10; %Numero de imagens a ser inseridas no
treinamento da VM
num_col = 200; %Numero de colunas das imagens (PIXEL)
num_lin = 200; %Numero de linhas das imagens (PIXEL)

for k=1:num_image
    %Local da imagem e formato, lembrando que é necessário q
as imagens estejam nomeadas sequencialmente
    Im{k}=imread(strcat('C:\Users\Rômulo\Documents\imagens
medicas\Imagens de Teste\imagens galdino\','int2str(k),
'.png'));
    Im{k}=rgb2gray(Im{k});
    Im{1,k}=num2cell(Im{1,k});
end

tic
% O tamanho do vetor x vai depender de quantas imagens vão ser
inseridas no treinamento
% Se tiverem 11 imagens a ser inseridas o vetor vai ficar:
% x = [[1];[2];[3];[4];[5];[6];[7];[8];[9];[10];[11]];

x = [[1];[2];[3];[4];[5];[6];[7];[8];[9];[10]];

%Inicio da SVR
for linha=1:num_lin
    for coluna=1:num_col
        for imagem=1:num_image
            dados{imagem,:} = Im{1,imagem}{linha,coluna};
        end
        dados_uint = cell2mat(dados(:));
        dados_double = im2double(dados_uint);
        linha
        coluna
        mdl = fitrsvm(x,dados_double, 'KernelFunction', 'rbf');
        result{linha,coluna} = predict(mdl, (num_image+1));
    end
end
end
toc
%Final da SVR
result_mat = cell2mat(result);
image_final = uint8(255 * (result_mat));
% Local da imagem ideal
image_ideal =
rgb2gray(imread(strcat('C:\Users\andre\Documents\imagens
medicas\Imagens de teste\imagens galdino\12.png')));

subplot(1,2,1), subimage(image_final)

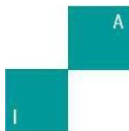
```

```
subplot(1,2,2), subimage(image_ideal)

% Calculo da MAE
image_ideal_mat = im2double(image_ideal);
coluna = 0;
linha = 0;
mae = 0;
aux = 0;
linha = 0;
coluna=0;
```

ARTIGO SUBMETIDO

Inteligencia Artificial 20(59) (2017), 123-127 doi: 10.4114/intartf.vol20iss59pp123-127



INTELIGENCIA ARTIFICIAL

<http://journal.iberamia.org>

**Development of a Hydrological Forecasting Model Using
Regression Support Vector Machine: an Computational Approach for Modelling of Arkansas River Basin**

Abstract: Investments in natural water resources are used for different purposes among the governmental bodies that manage these resources. Among these purposes are water supply, power generation, irrigation and navigation. The prediction of the accuracy of such natural resources can have a great impact and relevance for society, since the precise estimation of the future behavior of these resources can guide the authorities in the process of decision making, regarding the prediction and mitigation of possible disasters environmental and socio-environmental impacts of climate change caused by global warming, thus reducing the operational costs of measures to be taken. To better target investments in the prevention and mitigation of the effects caused by the lack of these resources, it is necessary to develop reliable forecasting methods. This research explores two methods of prediction, neural networks and regression tree. On the other hand this article proposes the implementation, application and analysis of the algorithm SVM (Support Vector Machine). Similarly, to the two previous ones, this proposed method estimates predictive models of images, aiming to consider future changes of water resources using information obtained from the set of NDVI historical data. These images show the features of the Arkansas River, which were obtained from the Space Research Center of the University of Texas, Austin. Finally, the accuracy of the three methods was compared with the aim of quantifying the robustness of the algorithms to predict fluid movements in the Arkansas river basin.

Resumo: Os investimentos em meios hídricos naturais são utilizados com finalidades distintas entre os órgãos governamentais que gerenciam esses recursos. Entre tais finalidades, destacam-se o abastecimento de água, a geração de energia, a irrigação e a navegação. A predição da acurácia de tais recursos naturais pode ter grande impacto e relevância para a sociedade, haja vista que a estimação precisa do comportamento futuro desses recursos pode orientar as autoridades no processo da tomada de decisão, no que concerne à previsão e mitigação de possíveis desastres ambientais e socioambientais, decorrentes de alterações climáticas produzidas pelo aquecimento global, reduzindo-se assim os custos operacionais das medidas a serem tomadas. Para melhor direcionar os investimentos na prevenção e na mitigação dos efeitos causados pela falta desses recursos, faz-se necessário o desenvolvimento de métodos confiáveis de previsão. Esta pesquisa explana dois métodos de previsão, redes neurais e árvore de regressão. Por outro lado este artigo propõe a implementação, aplicação e análise do algoritmo SVM (Support Vector Machine). Similarmente, aos dois anteriores, este método proposto, estima modelos predictivos de imagens, visando ponderar mudanças futuras dos recursos hídricos usando informações obtidas a partir do conjunto dos dados históricos NDVI. Estas imagens mostram as feições do rio Arkansas, que foram obtidas do Centro de Pesquisas Espaciais da University of Texas, Austin. Finalmente foi comparada a precisão dos três métodos, com o objetivo de quantificar a robustez dos algoritmos para prever movimentos de fluídos na bacia do rio Arkansas.

Keywords: Water Resources, Prediction, Neural Networks, Regression Tree, SVM, Remote Sensing.

Palavras Chaves: Recursos Hídricos, Predição, Redes Neurais, Árvore de Regressão, SVM, Sensoriamento Remoto.

1 Introdução

Com o crescimento da população e da atividade econômica, muitos países encontram-se em situação de escassez de recursos essenciais, como a água, ou estão enfrentando limites para seu desenvolvimento econômico.

O objetivo deste trabalho é avaliar o desempenho de três métodos de aprendizado de máquina, os quais sejam árvores de regressão, uma rede neural e o SVR, para estimar e prever o comportamento dinâmico dos recursos hídricos obtidos nos registros dos dados de uma série temporal de imagens de sensoriamento remoto NDVI (*Normal Difference Vegetation Index*). Através de séries temporais de imagens de sensoriamento remoto, pode-se fazer previsões para qualquer padrão observado nas feições da superfície do solo: queimadas, desertificação, recursos hídricos monitorada pelo sensor NDVI [1], [2]. A escolha de usar as imagens de sensoriamento remoto fornecidas pelo Centro Espacial da Universidade do Texas é oportuna, uma vez que essas imagens mostram uma dinâmica da bacia do rio Arkansas. Uma aplicação relevante deste estudo pode ser na previsão da escassez de recursos hídricos que é um problema de relevância mundial. Os recursos naturais podem ser grandemente afetados, seja por meio da ação humana (prática pecuária, prática da agricultura de sequeiro, pastagem e irrigação), seja pela ação da natureza (mudanças climáticas). Assim, torna-se possível analisar e avaliar como essa relação complexa, homem-natureza, e natureza-natureza pode chegar a impactar o meio ambiente[3].

Na literatura, a maioria dos casos com os algoritmos estudados (Rede Neural, Árvore de Regressão, SVR) são analisados a partir de grandes dados existentes. O presente trabalho foi elaborado a partir de recortes feitos do rio Arkansas no Texas, localizados nos Estados Unidos e, a partir dos dados coletados, fez-se uma análise que possibilitou verificar, dentre os três algoritmos estudados, qual deles apresenta melhor desempenho computacional em previsões com dados menores, de modo a viabilizar a resolução dos problemas hídricos em locais ínfimos com maior celeridade pelas autoridades competentes, trazendo mais agilidade na tomada de decisões no que concerne a diminuição/aumento e danos a esse recurso natural.

Os testes abordados neste estudo foram realizados usando sequências temporais de imagens de sensoriamento remoto NDVI, obtidas na superfície do solo do estado de Arkansas, nos Estados Unidos, o que permitiu idear quais as mudanças possíveis no conteúdo hídrico do rio Arkansas ao longo do tempo.

Nesse sentido, este trabalho de pesquisa é um novo desenrolar de maneira aprofundada de pesquisas já feitas e apresentadas no artigo GEOBIA [4].

O restante desse trabalho está organizado da seguinte maneira: Na seção 2, Material e Métodos, que faz uma breve revisão sobre os algoritmos de regressão utilizados. Foram realizados testes a partir de séries de imagens de sensoriamento remoto que refletiram a mudança no ambiente. Ademais, foi avaliada a acurácia da aplicação no domínio do espaço (pixel a pixel). Na seção3, Formulação propõe estratégias de predição de séries temporais de imagens (metodologia), respectivamente à amostra das métricas MAE, RMSE utilizadas para exploração e discussão dos resultados. Na seção 4, Experimentos e Resultados, serão mostrados os resultados dos testes junto às discussões a respeito dos resultados obtidos e as formas de implementação. As conclusões sobre este trabalho e perspectivas futuras são discutidas na secção Conclusões.

2 Material e Métodos

A análise estimativa para previsão de séries temporais a partir das sequencias de NDVI foi gerada a partir da aplicação de uma árvore de regressão, uma rede neural *Multi Layer Perceptron* e o SVR. As Máquinas de Vetores Suporte (*Support Vector Machines - SVMs*) com extensão SVR (Regressão por Vetores de Suporte) constitui numa técnica embasada na Teoria de Aprendizado Estatístico Vapnik e Chervonenkis (VC), que vem recebendo grande atenção nos últimos anos [5] [6].

Para tanto, foram realizadas previsões com redes neurais. As redes neurais são sistemas distribuídos paralelos, constituídas de unidades de processamento simples (neurônios artificiais) que calculam determinadas funções matemáticas (tipicamente não-linear). Tais unidades estão dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos, estas ligações estão associadas com pesos, que armazenam os conhecimentos adquiridos pelo modelo e servem para considerar a entrada recebida em cada rede de neurônios [7].

O neurônio artificial é uma forma de neurônio biológico de simulação, sendo a unidade básica para o processamento de informações em uma rede neural [8]. O modelo básico, representado pelo diagrama de blocos da Figura 1, mostra a relação entre os elementos do neurônio artificial:

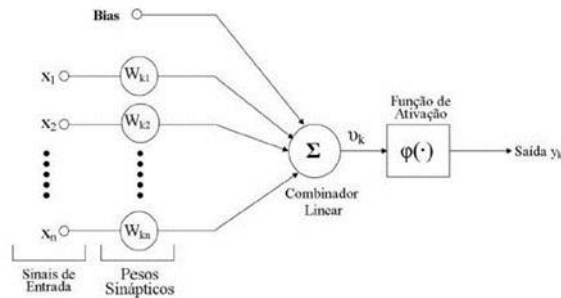


Figura 1. Modelo de um neurônio

Com base na Figura anterior, verifica-se que o neurônio tem os seguintes elementos:

- 1) sinais de entrada x_1, x_2, \dots, x_n ;
- 2) um conjunto de sinapses com os seus pesos sinápticos $w_{k1}, w_{k2}, \dots, w_{kn}$. Os pesos sinápticos são representados por w_{kj} , onde k representa o índice do neurônio e j refere-se ao terminal de entrada da sinapse ao qual o peso sináptico se refere;
- 3) Um combinador linear u , representado pela soma dos sinais de entrada ponderado por pesos sinápticos;
- 4) A função de ativação ϕ para restringir a extensão da saída de um neurônio;
- 5) Um bias b_k , que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo, respectivamente.

O algoritmo de treinamento utilizado neste estudo para a previsão de séries temporais foi o LevenbergMarquardt. Este algoritmo tem sido amplamente utilizado na literatura em problemas gerais que envolvam redes neurais, incluindo redes recorrentes e se encaixa no paradigma de aprendizagem supervisionada, visto que utiliza dados conhecidos como treinamento para prever resultados futuros [9].

$$W(n+1) = W(n) - (J^t(n) J(n) + \mu I)^{-1} J^T(n) e(W) \quad (1)$$

Na equação, J é a matriz jacobiana das derivadas dos erros relacionados aos pesos, μ é um valor escalar, e (W) é um vetor de erros. Se o valor de μ é grande, a expressão acima representa uma descida mais rápida, para pequenos valores de μ a expressão é reduzida para o método de Gauss Newton, que é mais rápido e mais preciso, próximo do erro mínimo. No caso em estudo, o treinamento foi feito com o ajuste dos parâmetros livres da rede, adaptando-o gradativamente ao ambiente externo.

A estimação e previsão de séries temporais de imagens 2D é um processo que depende de um grande conjunto de probabilidades, onde regiões (conglomerados de pixels) podem crescer e se deslocar para quaisquer direções [10]. Assim, as propostas de estratégias de previsão de sequências de imagens são expressas sem que isso acarrete num problema de dimensionalidade.

A previsão no domínio pixel é realizada quando cada pixel da imagem da série temporal está como uma função da sua intensidade de pixel, de modo que a mudança de intensidade de pixel é ajustada pela função de saída da rede neural. Este é o método de previsão direta, sem qualquer transformação matemática.

Outra abordagem interessante de regressão é o Algoritmo árvore de regressão, que é uma variação da árvore de decisão, desenvolvido por Leo Breiman nos anos 80, esses tipos de algoritmos por indução são muito idênticos. A estrutura de uma árvore de regressão (denominado na literatura, Regression Tree - RT) é uma variação do algoritmo CART. Contudo, é semelhante a uma árvore de decisão, pois também é formada por conjuntos de nós de decisão, que diferem no resultado, que em vez de classificar fornece previsões [11] [12].

No problema de complexidade é feita a divisão em problemas mais simples, a qual utiliza o conceito de repetição para aplicar a mesma técnica. Todas as soluções geradas dos subproblemas podem ser combinadas, em formato de árvore, para alcançar uma solução em problemas complexos [12].

Nesse estudo, a árvore de regressão tem como entradas os pixels do histórico das imagens da série temporal, resultando numa saída do pixel no tempo futuro. O procedimento é repetido para cada pixel, até varrer as imagens completamente. Além disso, existe uma forte compressão das sequências de imagens, o que reduz redundâncias e concentra a maior parte da informação dentro de poucos coeficientes[12],[13].

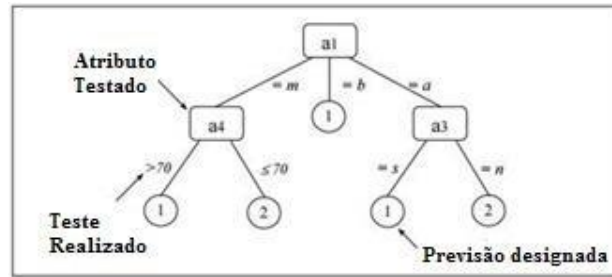


Figura 2. Estrutura de uma Árvore de Regressão

No exemplo da Figura 2, a1, a3 e a4 representam os atributos distribuídos na árvore de acordo com seu nível informativo. Nos arcos, são testados os valores dos atributos designados ao nó a que pertencem. Valores categóricos, são representados por uma igualdade, por exemplo $= m$, onde m é um valor do atributo. Para valores quantitativos, é estrado intervalo de valor, por exemplo >70 , sendo este intervalo obtido através de cálculo. Cada círculo, ao final dos ramos da árvore, indica a previsão associada aos nós das folhas. A regressão ocorre ao se percorrer o caminho que se inicia no nodo raiz ($a1$) e se estende até as folhas. Para a regressão, os valores das folhas são quantitativos [13].

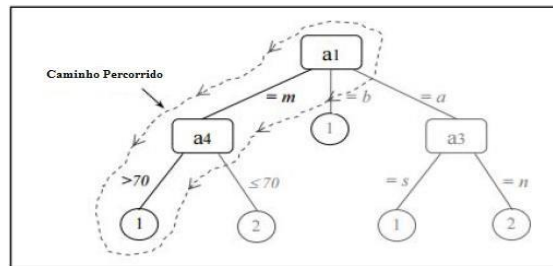


Figura 3. Exemplo de um caminho percorrido

O caminho destacado na Figura 3 indica um caminho para resultados de regressão. Através do caminho descrito por uma árvore é possível derivar regras. As árvores e as regras são geralmente utilizadas em conjunto. De acordo com a aplicação, as árvores tendem a crescer muito, sendo elas substituídas pelas regras.

A árvore de regressão é uma ferramenta de decisão cuja estrutura treinada pode revelar um conjunto de regras explanáveis e com certa precisão. O problema de regressão consiste na criação de um modelo funcional que relacione o verdadeiro valor da saída com os regressores (variáveis de entrada). Ao aplicar esta técnica, assume-se a variável de saída a emular uma previsão, utilizando-se para a resolução de problemas de regressão. O processo de treino de uma árvore de regressão consiste num algoritmo de partição do conjunto de treino, em cada nó da árvore, em dois subconjuntos disjuntos, tratando-se, então, de uma árvore binária onde esse ciclo binário fará a divisão sucessivas de possíveis soluções. Assim, a árvore é criada com base em regras de parada que interferem e contribuem na melhoria da previsão pela árvore [14].

Nas regras de divisão, existe um conceito fundamental chamado de ganho de informações, o qual utiliza o conceito de entropia, que tem como base medir a aleatoriedade de uma variável aleatória. A árvore de regressão usa a entropia para medir a aleatoriedade (complexabilidade em fazer previsões) na variável alvo (meta) ou observações rotuladas. Em todos os nós de regressão o atributo que mais reduz a aleatoriedade da variável meta é nomeado para dividir os dados. O ganho de informações mede a redução da entropia para cada atributo, a partir das partições adquirido de acordo com os valores do atributo. A construção de uma árvore de regressão é guiada pelo propósito de minimizar a entropia, que refere-se a complexidade em fazer previsões da variável alvo (meta) [15]. Normalmente, na entropia de uma variável aleatória v com valores $v = [v_1, v_2, v_3, \dots, v_k]$, o tipo de medida da entropia é *bits*, utilizando o logaritmo na base 2, cada um com probabilidade $P(v_k)$, definida pela equação:

$$H(v) = \sum_k P(v_k) \log_2 \frac{1}{P(v_k)} = - \sum_k P(v_k) \log_2 P(v_k) \quad (2)$$

O ganho de informações dos testes aplicados no atributo pode ser visto na equação abaixo, que representa a redução prevista na entropia:

$$\text{Ganho} = B\left(\frac{p}{p+n}\right) - \sum_{k=1}^d \frac{p_k + n_k}{p+n} B\left(\frac{p_k}{p_k + n_k}\right) \quad (3)$$

Na fórmula acima, p representa os exemplos positivos e n exemplos negativos de entrada. Para cada subconjunto tem p_k que representa os exemplos positivos e n_k os exemplos negativos de entrada. Se continuar ao longo dessa ramificação, precisamos de $B(p_k/(p_k + n_k))$, *bits* para complementar informações para responder à interpretação do problema.

No problema de regressão, a função de custo que reduz é, geralmente, o erro quadrático. Mediante disso, a constante relacionada às folhas de uma árvore de regressão é a média do atributo alvo (meta) das amostras de treinamento que caem da folha. O ganho de informação nos faz entender o seguinte procedimento que construir uma árvore de regressão, é basear-se na árvore de decisão e praticamente todos os ganhos de informação são semelhantes, o que diferencia é a função de custo, ou seja, enquanto a árvore de regressão minimiza os custos, a árvore de decisão maximiza esses custos na otimização. É uma parte da matemática que inclui características de resultados para formar funções que devem ter suas restrições e que podem ser impostas [15] [16].

Outro algoritmo de aprendizado de máquina é o SVM no contexto de regressão é nomeado de Regressão por Vetores de Suporte (SVR). Nesse contexto, o problema básico de regressão é encontrar uma função a qual aproxima determinado conjunto de dados. Nesse caso, deseja-se encontrar uma função $f(x)$ que aproxima um determinado vetor y a menos de um erro de previsão ε especificado [16]. Existem diversas técnicas de aprendizagem de máquina para solução de problemas de regressão que, ao serem empregadas para solucionar o mesmo problema, essas estratégias podem alcançar resultados diferentes. Este trabalho optou por utilizar um método de regressão por vetores de suporte – SVR, tendo em vista que esse método tem apresentado bons resultados em estudos comparativos de técnicas de previsão de séries temporais, aplicados a diferentes áreas do conhecimento [17] [18].

O funcionamento do algoritmo Regressão por Vetores de Suporte (SVR) está baseado na metodologia do aprendizado supervisionado. Tal metodologia faz com que um sistema aprenda uma função a partir de uma amostra de dados de treinamento, onde x denomina o espaço de entrada e y o espaço de saída. Nesse sentido, o objetivo do SVR é encontrar uma função que apresente, no máximo, um desvio ε em relação aos valores alvo, [16] [19].

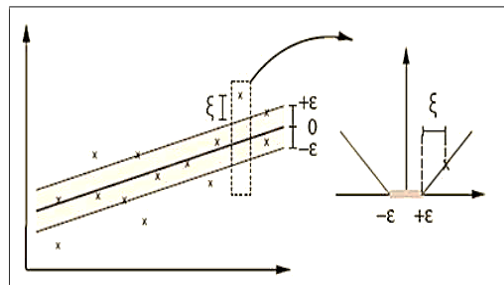


Figura 4. Função de aproximação SVR (lado esquerdo), Gráfico função de perda insensível (lado direito), [20] adaptado

Assim na Figura 4 observa-se que para valores entre $-\varepsilon + \varepsilon$ que formam a margem, onde não existe penalização de resultados [20] [21], ou seja, apenas os valores fora da margem são penalizados, neste caso, de forma linear. Isso corresponde a lidar com a chamada função de perda, dada por:

$$|\varepsilon|_{\varepsilon} = \begin{cases} 0, & \text{se } |\varepsilon| \leq \varepsilon \\ |\varepsilon| - \varepsilon, & \text{caso contrário} \end{cases}$$

A regressão de vetores de suporte SVR (Suporte por Vetores de Regressão) é construído e fundamentado em uma máquina de vetor de suporte SVM (Suporte por Vetores de Máquina). O modelo produzido por SVR depende apenas de um subconjunto dos dados de formação e por sua função de custo, que é utilizada para a construção do modelo, ignorando os dados de formação perto do modelo de previsão. Se uma amostra de dados de treinamento $\{(x_1, y_1), \dots, (x_L, y_L)\}$, onde x indica o espaço de padrões de entrada, sendo o objetivo de encontrar uma função $f(x)$ que apresente no máximo um desvio ε dos alvos y , obtido para todos os dados de treinamento. Logo, descreve-se a função linear de aproximação por $f(x) = \langle w, x \rangle + b$ com $w \in x$, b onde $\langle w, x \rangle$ denota o produto interno em x . Contudo, nem sempre é possível garantir a viabilidade do problema, já que existem pontos que violam as restrições, chegando-se ou resultando em previsões [16] [20] [22]. Assim, tem-se a seguinte equação:

$$\begin{aligned} & \text{Minimize: } \frac{1}{2} \|w\|^2 \\ & \text{Sujeito: } y_i - (w \cdot x_i) - b \leq \varepsilon \quad (4) \end{aligned}$$

$$(w \cdot x_i) + b - y_i \leq \varepsilon, \forall i = 1 \dots L$$

O algoritmo SVR fornece como alternativa trabalhar em um espaço de alta dimensionalidade. Assim, pode-se realizar um mapeamento dos dados de entrada para um espaço de dimensão maior, onde a regressão linear torna-se possível. Para isso, utiliza-se a abordagem baseada em funções Kernel.

Dentre as funções Kernel mais utilizadas no algoritmo SVR destacam-se o Kernel: linear, polinomial homogêneo, polinomial não homogêneo, sigmoidal e RBF [23]. No nosso caso de estudo, utilizamos o RBF (Funções de Base Radial), descrito pela seguinte equação:

$$K(x, x') = e^{-\gamma \|x - x'\|^2} \quad (5)$$

Ademais, importante ressaltar que um kernel expressa uma medida de similaridade entre vetores. O kernel RBF representa esta semelhança como uma função decadente da distância entre os vetores (a norma quadrada de sua distância). Ou seja, se os dois vetores estiverem próximos junto $\|x - x'\|$ os resultados serão pequenos. Então, enquanto $\gamma > 0$, segue-se que $-\gamma \|x - x'\|^2$ será maior. Assim, vetores mais próximos têm um valor de kernel RBF maior que vetores mais distantes. Esta função tem a forma de uma curva em forma de hipérbole. O parâmetro γ define a largura da curva em forma de hipérbole: quanto maior o valor de γ mais estreita será a hipérbole; da mesma forma, pequenos valores de γ produzem hipérbole largas [20] [23].

3 Formulação

Esta pesquisa apresenta três métodos de previsão de imagens de sensoriamento remoto da NASA, fornecidos pelo Centro de Pesquisas Espaciais da Universidade do Texas, registrados pelos satélites AQUA, com amostras de uma área que inclui o Estado do Arkansas e regiões vizinhas. As imagens têm dimensões de 5493x5340 pixels e uma codificação de 8 bits. Os dados começaram a ser capturados no ano 2002 no período de maio a dezembro, do satélite AQUA Data com o sensor MODIS.

O intervalo de tempo entre cada cena é de 16 dias, período que o satélite recobre o globo terrestre diariamente e, neste intervalo de dias, completa um ciclo de imageamento, compostos pelos pixels de radiométrica e geométrica [24]. Os valores de NDVI (Índice de Vegetação por Diferença Normalizada) são calculados pela seguinte equação:

$$NDVI = \frac{P_{NIR} - P_{RED}}{P_{NIR} + P_{RED}} \quad (6)$$

NDVI - Índice de vegetação por diferença normalizada;

P_{NIR} - Refletância de superfície no comprimento de onda do infravermelho; P_{RED}

- Refletância de superfície no comprimento de onda vermelho.

As imagens produzidas a seguir são resultados dessa equação, com estudo em escala de cinza. Devido às dimensões muito grandes das imagens brutas originais, foi necessário escolher uma área de estudo que mostrasse a variação dos recursos de água para ser cortada e processada mais tarde. A área escolhida foi correspondente à cidade de Fort Smith, que é cortada pelo Arkansas River [25].

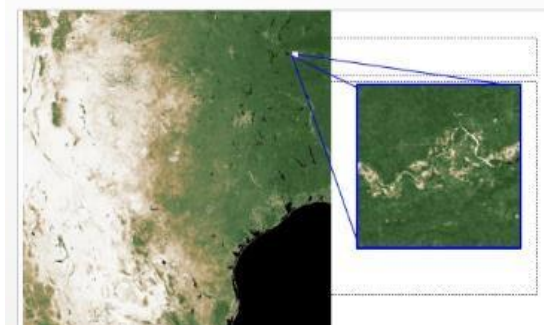


Figura 5. Área escolhida para análise

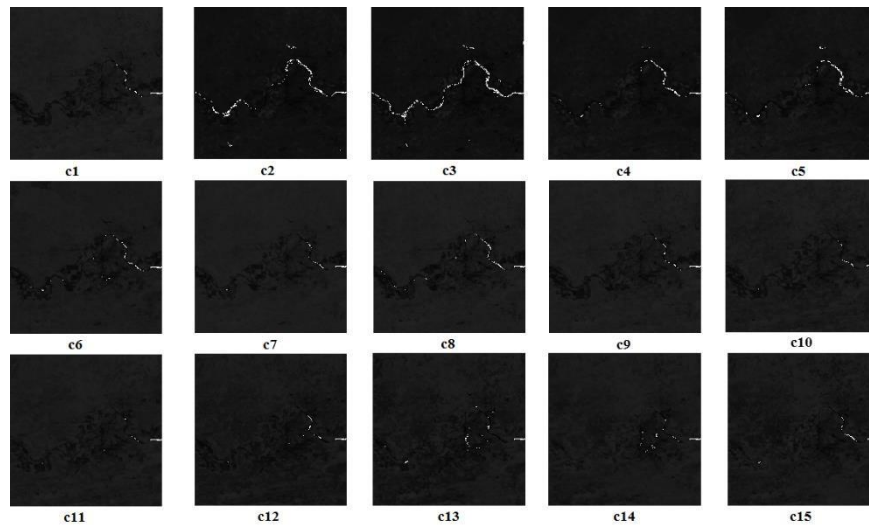


Figura 6. Série temporal de imagens recortada do Rio Arkansas.

Este trabalho de pesquisa foi desenvolvido com o apoio do seguinte software *Matlab* (software amplamente utilizado para pesquisa), desenvolvimento e análise de dados envolvendo métodos computacionais [26] e o software *QGIS*, cuja função é mostrar uma melhoria visual em escala de cor e o percentual da área do rio. Neste trabalho, o *Matlab* foi usado para implementar o SVR, Árvore de Regressão e a Rede Neural.

O método utilizado para validação e acurácia dos três algoritmos foi a métrica MAE (*Medium Absolute Error*) e o RMSE (*root-mean-square-error*). A justificativa para utilização dessas medidas é a necessidade de se medir a qualidade dos modelos e também o poder de previsibilidade do modelo. Um erro é definido como a diferença entre o valor real e o valor previsto, portanto, quanto menores são os erros, melhor é o modelo e a capacidade de prever valores futuros [26] [27]. O MAE e o RMSE podem ser usados em conjunto para diagnosticar a variação nos erros em um conjunto de previsões. No nosso caso de estudo, este foi aplicado sobre os pixels das imagens real e estimada, segundo as equações (7) (8). O erro MAE é uma métrica de erro estatístico usado para medir a grandeza da proximidade entre o resultado previsto e o real. Quanto mais próximo o valor da métrica de zero, mais eficaz é o método de previsão [27] [31]. Nas equações seguintes, xi deve ser sub-índice corresponde ao valor real, yi corresponde ao valor previsto e N é o número de repetições.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (7)$$

O RMSE é uma regra de pontuação quadrática que também mede a média do erro, ou seja, ele penaliza erros grandes e também é uma métrica de erro estatístico empregado para medir a é uma medida utilizada frequentemente as diferenças entre os valores (amostra) previstos por um modelo ou um estimador e os valores observados, sendo que é a raiz quadrada da média das diferenças quadradas entre previsão e observação real [28] [29], calculado pela seguinte equação:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (8)$$

Das 15 observações da série analisada, 05 foram reservadas para teste com valores previstos pela rede neural, árvore de regressão e o svr.

4 Experimentos e Resultados

A técnica proposta foi validada através de uma série de experimentos. Primeira experiência: as imagens abaixo mostram as previsões feitas em metodologia de previsão aplicada às sequências 11, 12, 13, 14,15 de imagens de sensoriamento remoto:

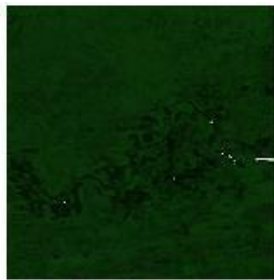
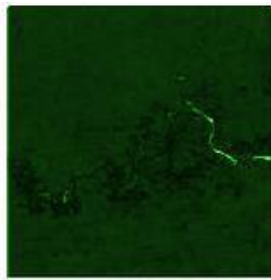
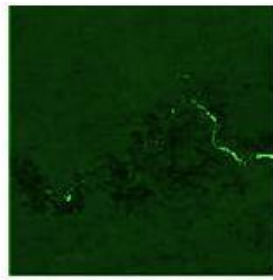


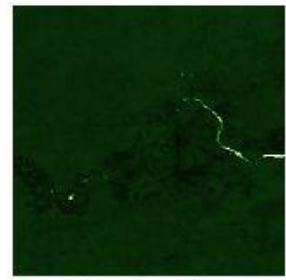
Imagem Real 11



**Espaço - Rede
Neural**



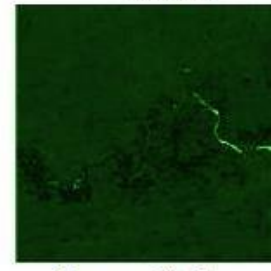
**Espaço - Árvore de
Regressão**



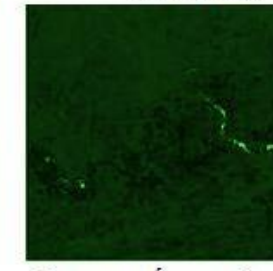
Espaço - SVR



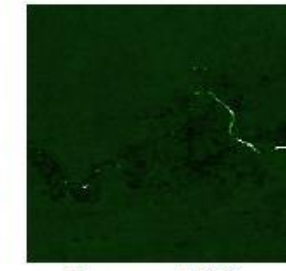
Imagem Real 12



**Espaço - Rede
Neural**



**Espaço - Árvore de
Regressão**



Espaço - SVR

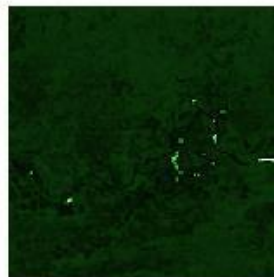
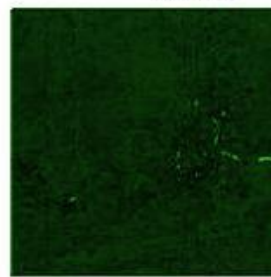
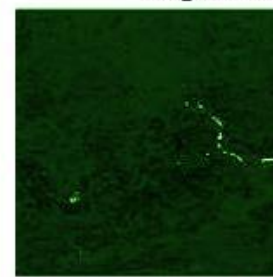


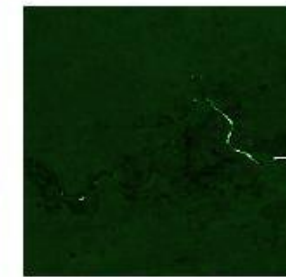
Imagem Real 13



**Espaço - Rede
Neural**



**Espaço - Árvore de
Regressão**



Espaço - SVR

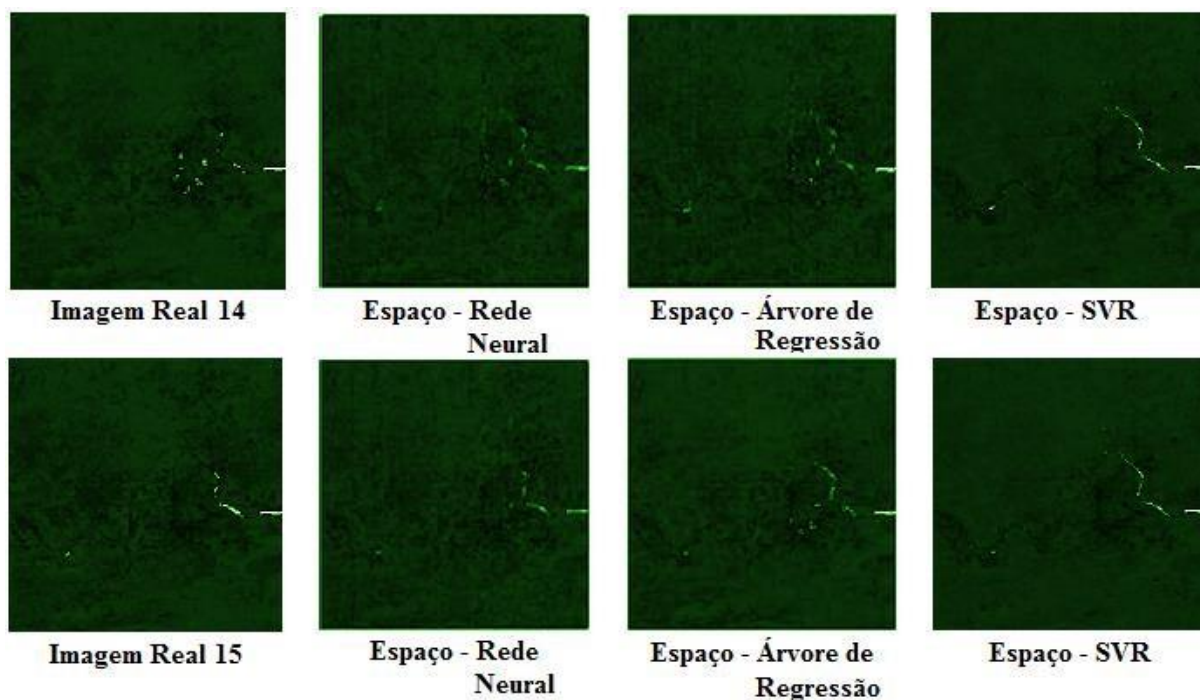


Figura 7. Imagens brutas resultante da previsão pixels para cada algoritmo, seguindo uma sequência da Imagem Real (ideal) e seus respectivos resultados no domínio de espaço, cadeia 11^a, 12^a, 13^a, 14^a, 15^a. Simulação de domínio de espaço, imagens foram feitas em níveis de resolução com fundo de cor verde para realce das feições, feitas pelo software QGIS.

No resultado das imagens, como primeiro experimento foi utilizada a metodologia aplicada a problemas de séries temporais, a partir da utilização de um sequenciamento de 10 imagens reais para fazer o treinamento do primeiro teste, resultando na 11^a prevista pelo algoritmo SVR e demais algoritmo (árvore de regressão, rede neural) e assim sucessivamente, resultados em 5 testes feitos. Útil cálculos de decrescente de erros de previsão, que são procedimentos adequados para o sucesso da predição com a utilização dos algoritmos é conveniente para obter melhores resultados de desempenho a um custo aceitável.

As imagens acima mostram que os resultados obtidos com os algoritmos e o uso das imagens em NDVI é vantajoso, visto que essas imagens por ser proporcional à biomassa existente são mais indicadas quando pretendese fazer comparações ao longo do tempo em uma determinada área de estudo. O resultado dessa razão ou equação é uma imagem monocromática em escala de cinza, com valores entre 1 e -1 [30], imagens resultantes do satélite AQUA, mostradas na figura 6. Quanto mais próximo de 1, indicará a presença de alta concentração de biomassa, quanto mais próximo de -1 mostrará a ausência de biomassa. Essas imagens monocromáticas foram coloridas com o software QGIS para melhor visualização dos resultados.

Como o caso em estudo refere-se ao Arkansas River, mostrou-se que a modificações nestas áreas de biomassa ao longo do tempo podem afetar de maneira drástica o percurso e existência do rio, as modificações provocadas pela própria natureza e pela ação humana a essa biomassa podem afetar diretamente o rio que contém um recurso renovável e tão precioso que é a água, já que a biomassa existe em conjunto. A decomposição das imagens implica em uma análise sob diferentes perspectivas, podendo-se atenuar aos erros de predição.

O uso das imagens em NDVI é relevante fazer estes levantamentos, com o uso de algoritmos que provam uma nova abordagem baseada em aprendizado de máquina e com intuito de representar de forma semelhante o comportamento real e previsto. Os resultados obtidos neste trabalho demonstram que tal abordagem permite alcançar uma representação satisfatória da disposição da imagem real e estimular novos estudos a fim de comparar o custo computacional da abordagem proposta com outras existentes.

Com isso, a razão para usarmos medidas é que precisamos medir a qualidade dos modelos e também o poder de previsibilidade do modelo. A escolha da métrica é determinada na prática dependendo de qual é o objetivo, ou seja, como deseja usar este modelo e suas restrições técnicas, matematicamente ou computacionalmente, no nosso caso seriam as predições de imagens de sensoriamento remoto em NDVI e as métricas utilizadas foram MAE e RMSE. O MAE (valor Absoluto Médio) é um valor absoluto e menos sensível a valores extremos, em comparação

com a função quadrada. Minimizar o MAE é encontrar uma aproximação para o valor mediano do resultado condicional da próxima observação (prevista) dadas as variáveis explicativas (o passado), em vez da expectativa condicional [16]. Já o uso do RMSE (Raiz do Valor do Quadrático Médio) minimizando essa métrica tenta encontrar uma aproximação para o valor esperado condicional da próxima observação (a ser prevista) dadas as variáveis explicativas (o passado na série temporal) [31], resumindo uma medida fornece a mediana e a outra a média. Resumindo encontrar a média e mediana dos resultados para estudos futuros dos seus resultados [31].

O MAE e o RMSE podem ser usados juntos para diagnosticar a variação nos erros em um conjunto de previsões. Tanto o MAE quanto o RMSE podem variar de 0 a ∞ . Uma vez que eles são pontuados com orientação negativa, valores menores são melhores. [16], [31].

A seguir mostraremos os resultados obtidos com uso dos três algoritmos (Rede Neural, Árvore de Regressão, SVR) com as métricas MAE e RMSE. As tabelas a seguir mostram o desempenho da Rede Neural, Árvore de Regressão e o SVR com estratégias de cálculo, feitas por meio do cálculo do erro MAE e RMSE.

Tabela 1 – ERRO MAE (%) DAS PREDIÇÕES (PIXEL Vs PIXEL)

Predições no Espaço feitas com a Rede Neural, Árvore de Regressão e o SVR			
Imagem	MAE (%) RN do Espaço	MAE (%) AR do Espaço	MAE (%) SVR do Espaço
Imagem Prevista 11	3,89	1,39	1,39
Imagem Prevista 12	4,27	1,77	1,09
Imagem Prevista 13	4,29	1,61	1,33
Imagem Prevista 14	4,23	1,31	1,25
Imagem Prevista 15	4,10	1,26	1,02

Tabela 2 – RMSE (%) DAS PREDIÇÕES (PIXEL Vs PIXEL)

Predições no Espaço feitas com a Rede Neural, Árvore de Regressão e o SVR			
Imagem	RMSE (%) RN do Espaço	RMSE (%) AR do Espaço	RMSE (%) SVR do Espaço
Imagem Prevista 11	1,739	0,621	0,621
Imagem Prevista 12	1,909	0,791	0,487
Imagem Prevista 13	1,918	0,720	0,594
Imagem Prevista 14	1,891	0,585	0,559
Imagem Prevista 15	1,833	0,563	0,456

Na Tabela 1, tanto árvore de regressão como o SVR, na métrica MAE, apresentaram uma acurácia melhor de maneira individual para cada teste, quando o treinamento e regressão das sequências NDVI foi estimada no domínio espacial pixel-pixel (x,y) em relação a rede neural. O MAE indicou o afastamento (desvio) mediano absoluto dos valores previstos em relação aos valores observados das imagens numa varredura de pixels. Em síntese, existe uma pontuação linear satisfatória dos resultados de previsões do modelo real para o previsto. O ideal seria que os valores do MAE, fossem o mais próximo de zero.

Na Tabela 2, os resultados do RMSE foram feitos de maneira a se obter a média entre os resultados dos testes para cada algoritmo, visto que essa métrica por conter pesos (quantidade de teste) é relativamente boa em relação a grandes erros. Isso significa que o RMSE é útil quando se deseja ter uma precisão do modelo de previsão, evitando assim erros grandes, particularmente indesejáveis. Também como o MAE, o ideal seria que os valores dos resultados fossem próximo de zero, porém, não descartando a eficácia para cada algoritmo utilizado neste trabalho. As métricas utilizadas neste trabalho mostraram a eficácia do desempenho dos algoritmos em fazer

predições, cada medida têm vantagens e desvantagens: a RMSE é melhor para mostrar desvios com atribuição da media dos testes; a MAE tem interpretação mais fácil ou individual para cada teste para se obter a mediana dos resultados.

Em relação aos algoritmos utilizados é mostrado um melhor desempenho em resultados pelo SVR em relação as outras técnicas de regressão, visto que as imagens mesmo com menor resolução espacial são melhores para observação de mudanças ao redor do rio e do próprio rio.

As imagens das sequências de treinamento, obtidas da região do Texas foram registradas pelo satélite “AQUA” da NASA, as feições dos corpos de água presentes na superfície da terra. Sendo assim, outras feições presentes ali aparecem pouco realçadas nas imagens registradas pelo satélite. Os corpos de água ai presentes são restritos ao rio Arkansas e representam uma pequena porcentagem dos objetos presentes nas sequencias das imagens NDVI.

Nos gráficos abaixo, são apresentados os resultados da métrica MAE RMSE, usadas para estimar o erro de previsão dos algoritmos.

Gráfico 1 - RESULTADOS DA PREVISÃO PIXEL COM A MÉTRICA (MAE) DA REDE NEURAL, ÁRVORE DE REGRESSÃO E O SVR

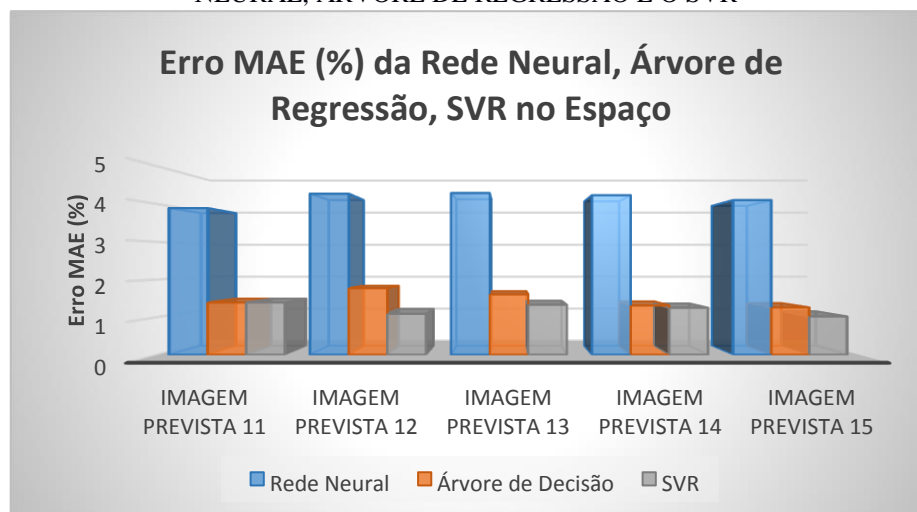
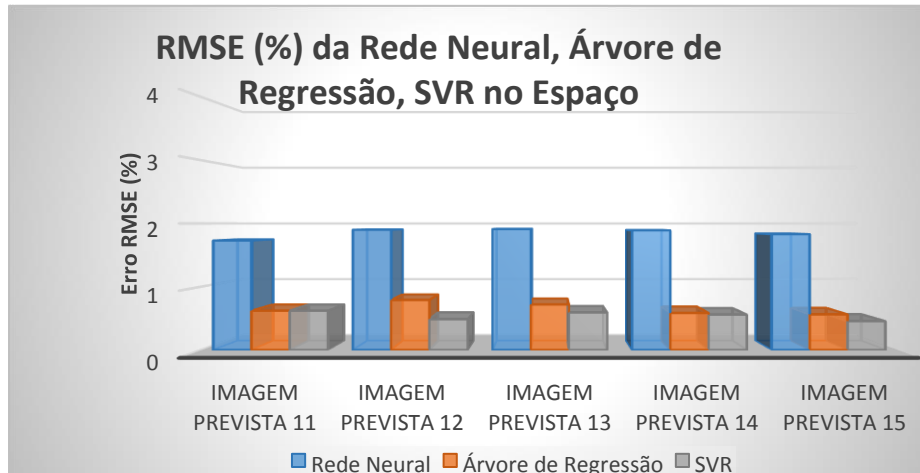


Gráfico 2 - RESULTADOS DA PREVISÃO PIXEL COM A MÉTRICA (RMSE) DA REDE NEURAL E ÁRVORE DE REGRESSÃO E O SVR



Os gráficos mostram o percentual dos erros MAE e RMSE das imagens da serie temporal do rio Arkansas. Considera-se então adequado validar os algoritmos com este tipo de métricas de avaliação MAE e RMSE, haja vista que os pixels de intensidade e significativos, presentes nas sequências NDVI, estão relacionados ao no nosso estudo de caso com o rio Arkansas. Outras feições dos objetos presentes nestas sequências, que não sejam água, têm resposta espectral do satélite AQUA visíveis, e são mostradas nas imagens. A radiação térmica emitida a partir destas regiões da superfície do solo são geradas e mostradas de dados de intensidade, nos registros das imagens capturados pelo sensor do satélite, tendo assim peso na estimação da métrica global do erro de predição MAE e RMSE.

A outra maneira de avaliar a diferença entre a imagem real e a previsão do rio Arkansas foi pelo calculo de porcentagem da area dos pixels das imagens. Utilizamos uma binarização focando somente nas mudanças do rio, assim os pixels foram classificados em duas faixas de valores: Cinza(vegetação densa, vegetação rasa, solo exposto); branco (rio). A classificação foi feita utilizando o software Qgis para mapear as areas dos pixels com a mesma cor, depois foi feito um grafico com porcentagem dos niveis das faixas predominante dos pixels, mostrado na figura 8.

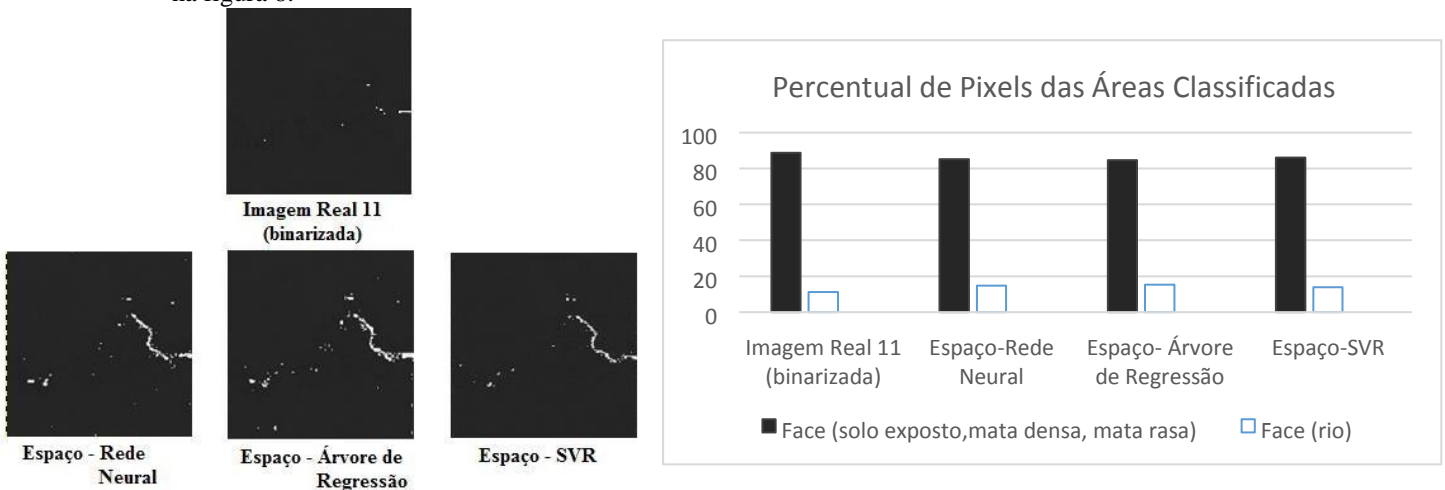


Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 11(binarizada)	88,74%	11,26%
Espaço- Rede Neural	85,17%	14,83%
Espaço- Árvore de Regressão	84,68%	15,32%
Espaço- SVR	86,11%	13,89%

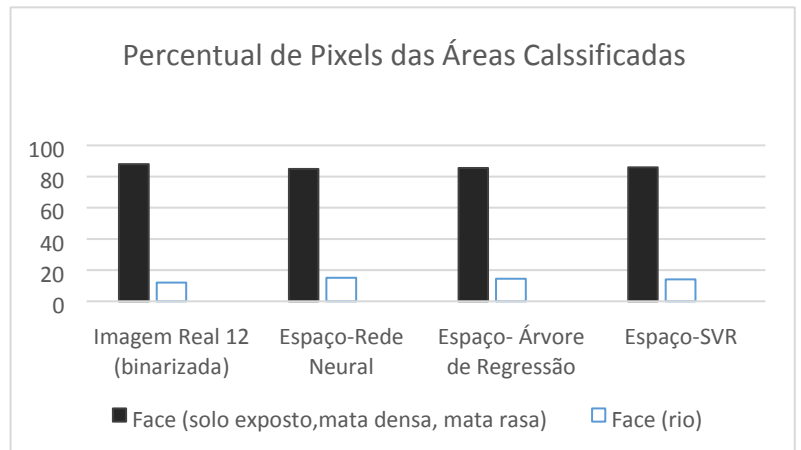
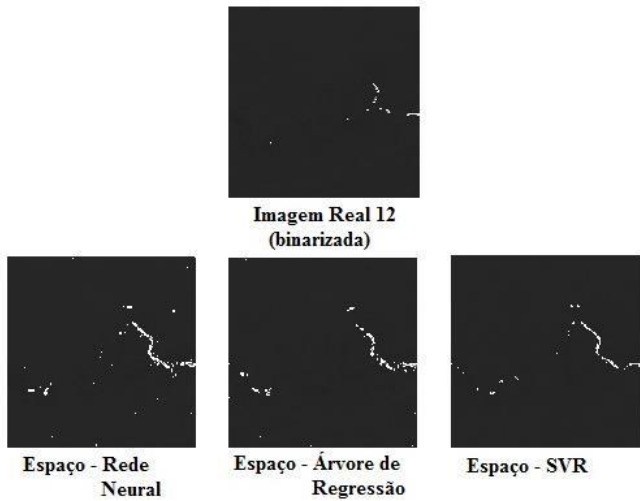


Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 12 (binarizada)	88,02%	11,98%
Espaço- Rede Neural	84,92%	15,08%
Espaço- Árvore de Regressão	85,46%	14,54%
Espaço- SVR	85,98%	14,02%

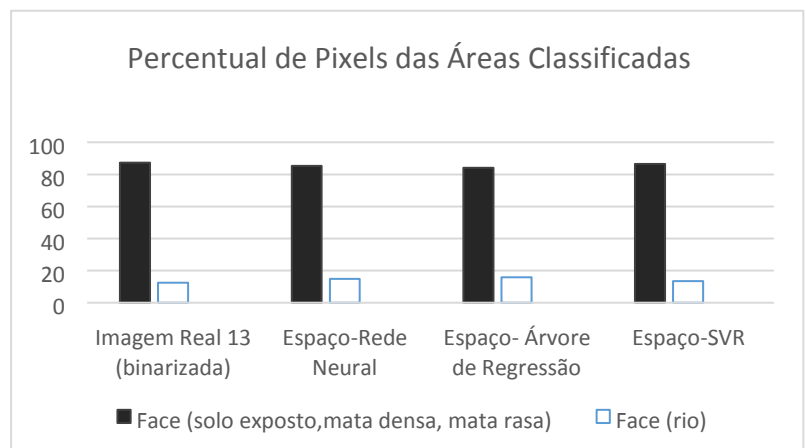
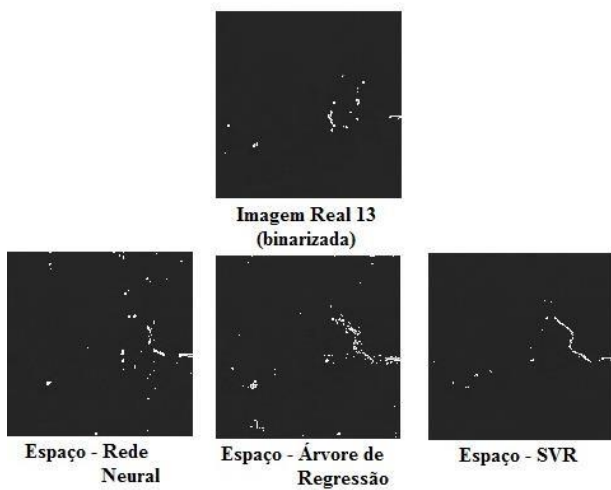


Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 13 (binarizada)	87,72%	12,38%
Espaço- Rede Neural	85,24%	14,76%
Espaço- Árvore de Regressão	84,08%	15,92%
Espaço- SVR	86,25%	13,75%

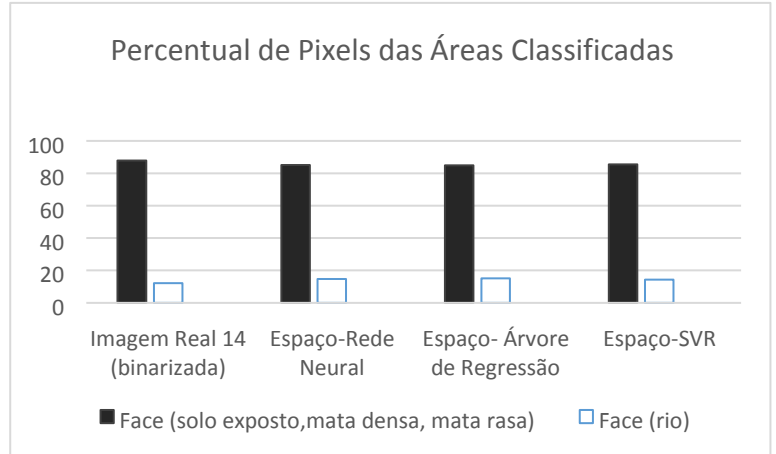
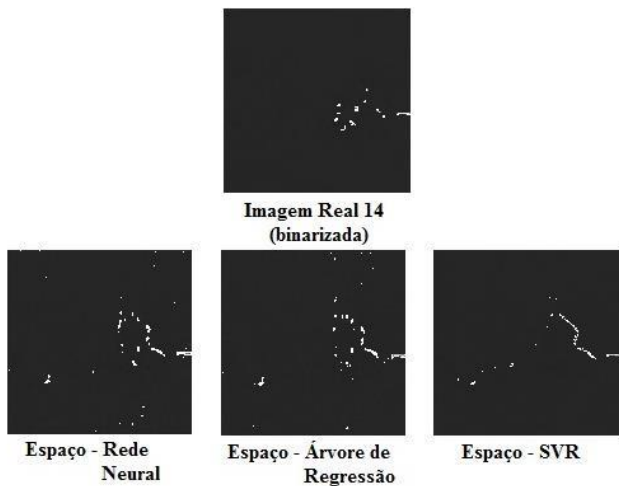


Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 14 (binarizada)	87,91%	12,09%
Espaço- Rede Neural	85,19%	14,81%
Espaço- Árvore de Regressão	84,93%	15,07%
Espaço- SVR	85,66%	14,34%

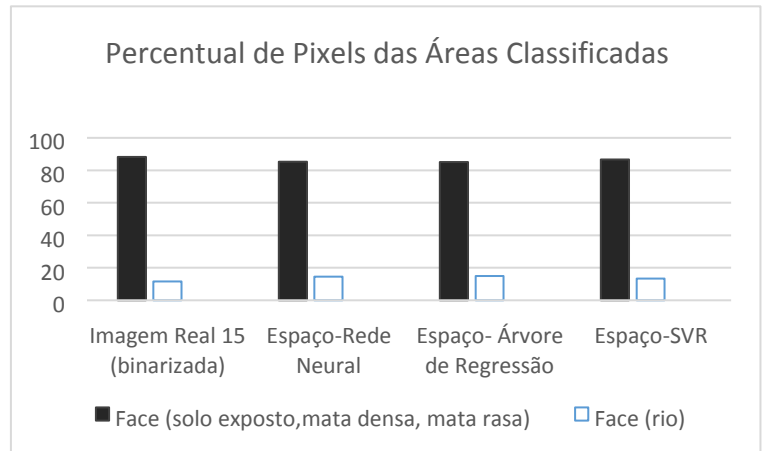
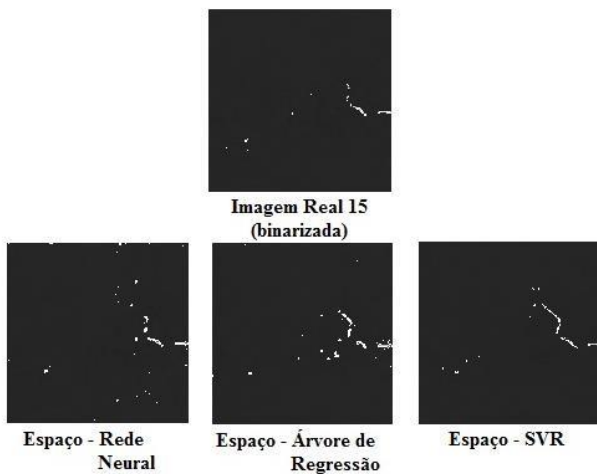


Imagem	Face (solo exposto, mata densa, mata rasa)	Face (rio)
Imagem Real 15 (binarizada)	88,33%	11,67%
Espaço- Rede Neural	85,38%	14,62%
Espaço- Árvore de Regressão	85,13%	14,87%
Espaço- SVR	86,58%	13,42%

Figura 8: mapas temáticos e porcentagem de pixels das áreas

Os testes foram executados com o mesmo tamanho da área escolhida e diferentes algoritmos de aprendizagem supervisionada, a fim de se fazer uma análise do comportamento do modelo e sua melhoria de previsão de acordo com as mudanças ocorridas. Assim com a segmentação binarizada, focando nas mudanças do rio Arkansas é possível modelar uma previsão significativa com o algoritmo SVR, como é mostrado pelo cálculo da porcentagem local, da área do rio, o método SVR, como é mostrado pelo cálculo da porcentagem local, da área do rio, o método

SVR permite estimar modelos de previsão com menores diferenças entre a imagem real de referência e o modelo numérico previsto pelo algoritmo

Comparando os resultados obtidos com a métrica percentual local é possível observar a acurácia dos resultados obtidos pelo algoritmo SVR com relação aos outros dois. Mesmo com o quantitativo que mostra uma melhor previsão da SVR, observa-se também nas sequências de imagens NDVI, um comportamento quase-sazonal que é típico em determinados tipos de séries temporais, como por exemplo as séries temporais financeiras e hidrológicas, as quais apresentam padrões e características que se repetem ao longo do tempo [32], [33].

Conclusão

O presente trabalho apresenta três algoritmos de aprendizado de máquina, que executam regressões preditivas, os quais foram utilizados para estimar modelos de previsão hidrológicos do rio Arkansas, a partir de séries temporais de sequências de imagens de sensoriamento remoto NDVI.

A ideia central deste trabalho foi implementar o método de aprendizado de máquina SVR para estimar um modelo de previsão hidrológica da bacia do rio Arkansas, usando sequências de imagens de satélite NDVI. Este método foi comparado e testada sua acurácia com relação a dois métodos concorrentes baseados numa rede neural Multicamada Perceptron e uma Arvore Decisão, visando construir um comitê de máquina, com o intuito de prever a probabilidade de surgimento de desastres ambientais. Desta forma, a tecnologia proposta neste artigo, tem potencial para auxiliar às autoridades públicas, que atuam na área de gestão ambiental, na tomada de decisão, no tocante à prevenção e mitigação de eventuais danos e desastres ambientais produzidos ao ecossistema pelo impacto da complexa interação homem-natureza/natureza-natureza. Assim, tal tecnologia constitui-se na prática, um sistema de alerta de risco ambiental.

As métricas de erro MAE e RMSE apresentaram resultados globais acurados, quando avaliadas sobre as imagens de saída estimadas dos algoritmos, no espaço de treinamento pixel-pixel neste estudo de caso. Os resultados obtidos com o SVR mostraram-se promissores para estimar modelos de previsão de séries temporais de imagens NDVI de sistemas complexos envolvendo interação homem-natureza e natureza-natureza.

Com a métrica local foi possível observar que a divisão dos algoritmos visa identificar padrões de comportamento da série e treinar múltiplos preditores especialistas nestes padrões temporais. Com a importância de se realizar a escolha dos preditores mais aptos em desempenho a realizarem a previsão da série e para a escolha dos dados para validar o treinamento e viabilizar resultados mais acurados.

Este trabalho abre novas perspectivas para a implementação de um comitê de máquinas maior que o atual, podendo envolver outras técnicas de aprendizado de máquina tais como: “deep learning-redes neurais de aprendizado profundo”, “redes probabilísticas de aprendizado bayesiana”, “Boosting,” entre outras. O uso de um comitê de aprendizado de máquina mais amplo seria motivado pela busca de se construir sistemas inteligentes que permitam a estimação e modelagem acurada do comportamento dinâmico dos sistemas complexos objetivo deste estudo, no caso, a bacia hídrica do rio Arkansas.

Num futuro próximo, este trabalho e o aperfeiçoamento das técnicas de previsão mencionadas acima, poderá levar à realização de novos casos de estudo, usando dados nativos obtidos do nordeste brasileiro, região que sofre com a escassez de água e que padece com processos ambientais/climáticos agressivos, tais como a desertificação e seca entre outros.

Agradecimentos

Agradecemos ao professor Byron Tapley, diretor chefe do Centro de Pesquisas Espaciais da University of Texas, nos Estados Unidos, pelo fornecimento dos dados experimentais que viabilizaram este trabalho de pesquisa, e à CAPES, pelo apoio financeiro na forma de bolsa de pesquisa dado a um discente recrutado para este projeto.

Referências

- [1] SOUSA, Willamys Rangel N.; SOLTO, M. V. S. ; MATOS, S. S. ; SILVA NETO, C. A.; DUARTE, C. R. . Extração automática de linhas de costa aplicada ao monitoramento de processos de erosão costeira. In: XVIII Simpósio Brasileiro de Sensoriamento Remoto, 2017, Santos. Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto - SBSR, 2017.
- [2] MOREIRA, M. Fundamentos do sensoriamento remoto e metodologias de aplicação: Universidade Federal de Viçosa, 2005.

- [3] CÂMARA, G.; DAVIS, Clodoveu (Org.) ; MONTEIRO, Antônio Miguel (Org.) ; PAIVA, J. A. (Org.) ; D'AGE, J. C. L. (Org.) . Geoprocessamento: Teoria e Aplicações. INPE - Edição On-line, 1999. v. 1.
- [4] SOUSA, W. R. N ; BANDEIRA, C. L. L. ; EWERTON C. RIBEIRO ; CASTRO, A. F. ; BUSTOS, H. I. A. ; SILVA, M. P. S. . Monitoring of desertification processes through trend estimates of time series. In: Geographic Object-Based Image Analysis - GEOBIA, 2012, Rio de Janeiro. Proceedings of the 4th GEOBIA, 2012. v. 1. p. 561-565.
- [5] BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006.
- [6] KWOK, J. T.-Y. Support vector mixture for classification and regression problems. In:IEEE. Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on. [S.l.], 1998. v. 1, p. 255–258.
- [7] MUKHOPADHYAY, A.; SHUKLA, B. P.; MUKHERJEE, D.; CHANDA, B. A novel neural network based meteorological image prediction from a given sequence of images In: IEEE. Emerging Applications of Information Technology (EAIT), 2011 Second international Conference on. [S.l.], 2011.
- [8] TAN, y.; BAI, B.; MOHAMMAD, M. Time series remote sensing based dynamic monitoring of land use and land cover change In: IEEE. Earth Observation and Remote Sensing Applications (EORSA), 2016 4th International Workshop on, 2016.
- [9] S. Haykin, *Neural networks: A comprehensive foundation.*, 2nd edition., Ed. Prentice Hall International, 2001.
- [10] AKBAR, H. et al. Training neural networks using Clonal Selection Algorithm and Particle Swarm Optimization: A comparisons for 3D object recognition. IEEE. Hybrid Intelligent Systems (HIS), 2011 11th International Conference on, 2012.
- [11] BHARGAVA, N; PUROHIT, R ; SHARMA, S. Prediction of arthritis using classification and regression tree algorithm. IEEE. Communication and Electronics Systems (ICCES), 2017 2nd International Conference. 2018.
- [12] PAN ,Z ,K; SONG, Y, Q; ZHOU, W. Classification of Urban Construction Land with Worldview-2 Remote Sensing Image Based on Classification and Regression Tree Algorithm. IEEE. Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference. 2017.
- [13] BERHAN, G. et al. Drought Information Mining from Satellite Images for Improved Climate Change Mitigation. IEEE. 2012 World Congress on Information and Communication Technologies. 2012.
- [14] MERINO, R, F; NUNEZ, M; Self-Adaptive Induction of Regression Trees. IEEE. IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 33, Issue: 8, Aug. 2011). 2011
- [15] NORVIG, P.; RUSSELL, S. *Inteligência Artificial, 3a Edição*. [S.l.]: Elsevier Brasil, 2014.
- [16] FACELI, K. et al. *Inteligência artificial: Uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC, v. 2, 2011.
- [17] YIGIT, M, E. et al. Genetic based approach to the synthesis of a cylindrical-rectangular microstrip conformal antenna using artificial neural network and Support Vector Regression models. IEE. Electrical and Electronics Engineering (ELECO), 2017 10th International Conference on, 2018.
- [18] FENTIS, A. et al. Short-term solar power forecasting using Support Vector Regression and feed-forward NN. IEEE. New Circuits and Systems Conference (NEWCAS), 2017 15th IEEE International, 2017.
- [19] FAN, J. et al. Vegetation coverage detection from very high resolution satellite imagery. IEEE. Visual Communications and Image Processing (VCIP), 2015. 2016.
- [20] SMOLA, A. J., & Schölkopf, B. A Tutorial on Support Vector Regression. Neuro COLT Technical Report NC-TR-98-030. 1998.
- [21] DATONG, L. et al. Fault prediction based on time series with online combined kernel SVR methods. IEEE. Instrumentation and Measurement Technology Conference, 2009. I2MTC '09. IEEE, 2009.
- [22] MISAGHI, S. et al. A hybrid model based on support vector regression and modified harmony search algorithm in time series prediction. IEEE. Fuzzy and Intelligent Systems (CFIS), 2017 5th Iranian Joint Congress on, 2017. [23] SHARMA, V; et al. An assessment of support vector machine kernel parameters using remotely sensed satellite data. IEEE. Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE International Conference on, 2017.
- [24] KHAN, J.; et al. Mapping MODIS LST NDVI Imagery for Drought Monitoring in Punjab Pakistan. IEEE. IEEE Access (Volume: 6), 2018.
- [25] MODIS NDVI data for Texas from CSR em: < <http://web.coral.tacc.utexas.edu/CSR/Public/NDVI-MODIS> >. Acesso em 20 de novembro de 2016.
- [26] ADAMCIK, F; BRÉDA, R; VAISPACHER, T. Analysis of GPS satellites position determination using MATLAB® environment. IEEE. Military Technologies (ICMT), 2015 International Conference on, 2015. [27] LIMA, C. M. M. “Um estudo sobre o método Mínimos Quadrados Móveis por Aproximações Iteradas,” Pontifícia Universidade Católica do Rio de Janeiro, 2009.
- [28] RODRIGUES, S, B. et al. combinação preditiva das técnicas arima e support vector regression (svr) na previsão do deslocamento do bloco h8 da usina hidrelétrica de itaipu com base nas séries temporais dos pêndulos. Cbdb.org.br, 2011.

- [29] ELJASSAR,M,M; HEYMAYED,E,E. Enhancing electric load forecasting of ARIMA and ANN using adaptive Fourier series. IEEE. Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual, 2017.
- [30] MENEZES,P;ALMEIDA,T. Introdução ao processamento de imagens de sensoriamento remote. UNB,EBook, Brasilia, 2012. Disponível em : < [http://www. cnpq. br/documents/10157/56b578c4 – 0fd5 – 4b9f – b82a – e9693e4f69d8](http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9f-b82a-e9693e4f69d8) > acesso em 28/05/2018.
- [31] A. DE MYTTENAERE, B. GOLDEN, B. LE GRAND, AND F. ROSSI, “Mean absolute percentage error for regression models,” *Neurocomputing*, vol. 192, 2016.
- [32] NI, H.; YIN, H. Exchange rate prediction using hybrid neural networks and trading indicators. *Neuro computing*, [S.l.], v.72, n.1315, p.2815–2823, Aug. 2009.
- [33] DABLEMONT, S. et al. Time series forecasting with SOM and local non-linear models -Application to the DAX30 index prediction. In: *WORKSHOP ON SELF-ORGANIZING MAPS*, Kitakyushu, Japan. Proceedings. [S.l.: s.n.], 2003