



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE - DOUTORADO

**NOVA ARQUITETURA UTILIZANDO REGRAS DE COMBINAÇÃO *FUZZY* PARA
MÉTODOS DE AGLOMERAÇÃO ESPACIAL APLICADA À EPIDEMIOLOGIA**

Laisa Ribeiro de Sá

João Pessoa/PB

2019

LAISA RIBEIRO DE SÁ

**NOVA ARQUITETURA UTILIZANDO REGRAS DE COMBINAÇÃO FUZZY PARA
MÉTODOS DE AGLOMERAÇÃO ESPACIAL APLICADA À EPIDEMIOLOGIA**

Tese apresentada ao Programa de Pós-Graduação em Modelos de Decisão e Saúde – Nível Doutorado, do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba, como requisito regulamentar para obtenção do título de doutora.

Linha de Pesquisa: Modelos de Decisão

Orientadores:

Prof. Dr. Ronei Marcos de Moraes

Prof. Dra. Jordana de Almeida Nogueira

João Pessoa/PB

2019

Catálogo na publicação
Seção de Catalogação e Classificação

S111n Sá, Laisa Ribeiro de.

NOVA ARQUITETURA UTILIZANDO REGRAS DE COMBINAÇÃO FUZZY
PARA MÉTODOS DE AGLOMERAÇÃO ESPACIAL APLICADA À
EPIDEMIOLOGIA / Laisa Ribeiro de Sá. - João Pessoa,
2019.

111 f. : il.

Orientação: Ronei Marcos de Moraes, Jordana de Almeida
Nogueira.

Tese (Doutorado) - UFPB/CCEN.

1. Métodos de aglomeração espacial. 2. Epidemiologia.
3. Conjuntos Fuzzy. 4. Regra de Votação Fuzzy. 5.
Combinação de classificadores. I. Moraes, Ronei Marcos
de. II. Nogueira, Jordana de Almeida. III. Título.

UFPB/BC

LAISA RIBEIRO DE SÁ

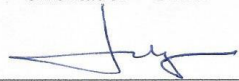
**NOVA ARQUITETURA UTILIZANDO REGRAS DE COMBINAÇÃO FUZZY PARA
MÉTODOS DE AGLOMERAÇÃO ESPACIAL APLICADA À EPIDEMIOLOGIA**

João Pessoa, 09 de Agosto de 2019

BANCA EXAMINADORA



Prof. Dr. Ronei Marcos de Moraes
Orientador – UFPB



Profa. Dra. Jordana de Almeida Nogueira
Orientadora – UFPB



Prof. Dr. Rodrigo Pinheiro de Toledo Vianna
Membro Interno – UFPB



Profa. Dra. Rafaela Queiroga Souto
Membro Externo – UFPB



Prof. Dr. Regivan Hugo Nunes Santiago
Membro Externo – UFRN

Dedico aos amores da minha vida, pedacinhos meus:

Ao meu primogênito, José Eduardo, que me acompanhou desde a barriga no mestrado e seguiu nas aulas do doutorado. Meu amor, foi você que me fez descobrir o quanto era forte e mesmo tão pequeno transformastes minha vida.

Ao meu pequeno, Pedro Henrique, que chegou no meio desta jornada. Foi um turbilhão de emoções, mas desde a barriga tu eras calma. Meu amor, seus sorrisos são tão doces, seu olhar tão sereno. Trouxeste paz.

AGRADECIMENTOS

Cheguei ao fim desta longa e árdua caminhada e gostaria de agradecer:

À Deus, em primeiro lugar, pelos dons concedidos para realizar este trabalho, dando a tranquilidade necessária para seguir em frente e não desanimar com as dificuldades.

Aos meus pais Gilson José Feliciano de Sá e Mariluce Ribeiro de Sá e a minha irmã Letícia Vitória, pelo constante apoio e enorme compreensão diante de todas as dificuldades, estando sempre presentes quando preciso e entendendo todas as minhas ausências.

Ao meu esposo Ricardo Moura Leite, que sempre esteve ao meu lado, dando forças para seguir em frente, e quando pensei em desistir deu-me coragem para prosseguir.

Aos meus filhos José Eduardo e Pedro Henrique que ensinaram o mais profundo e sincero amor. Quanto foi difícil para nós as ausências. Quantas aulas assistidas juntos. Quantas noites amamentando e estudando ao mesmo tempo. Quantos “você é louca” escutei por ter tido vocês nessa fase, mas mamãe conseguiu e foi por vocês.

A Fátima e Josivaldo por ter ajudado com os meninos, sem vocês não teria conseguido.

A família e amigos que deram todo o apoio nessa minha caminhada.

Agradeço as minhas amigas Camila, Rafaela e Manu, por compartilharem as mesmas loucuras, medos e sonhos.

Aos meus orientadores, Jordana de Almeida Nogueira e Ronei Marcos de Moraes pelos ensinamentos concedidos, pela paciência, competência, incentivo e confiança depositada.

A Professora Liliane Machado pela motivação constante.

À banca examinadora deste trabalho, pelo tempo dedicado e as contribuições para melhorias nessa pesquisa.

Aos integrantes do laboratório, Leapig/Labteve, que me auxiliaram em todos os momentos no desenvolvimento da pesquisa.

A CAPES pelo apoio financeiro, que foi fundamental para o andamento da pesquisa.

“A satisfação reside no esforço, não no resultado obtido. O esforço total é a plena vitória”.

Mahatma Gandhi

RESUMO

Na área da saúde, principalmente no âmbito da epidemiologia, necessita-se de métodos que consigam georreferenciar o evento estudado, visto que esta informação é primordial para o entendimento do fenômeno e para o processo de tomada de decisão. A combinação dos métodos de aglomeração espacial com a inclusão das incertezas presentes no processo saúde-doença, produz melhores resultados do que se utilizados individualmente, com intuito de produzir informações mais fidedignas e tomada de decisões mais precisas. O objetivo do estudo foi desenvolver uma nova arquitetura utilizando regras de combinação *fuzzy* para métodos de aglomeração espacial aplicada à epidemiologia. Trata-se de um estudo ecológico, retrospectivo de abordagem quantitativa. Para construção da nova arquitetura que foi testada em três estudos de casos distintos usando os dados epidemiológicos dos casos de Dengue, notificados no estado da Paraíba, Brasil, dos anos de 2011 e 2017. Os estudos de casos demonstraram a funcionalidade da arquitetura proposta através da combinação dos métodos utilizando diferentes regras de combinação, além da comparação com arquitetura anterior com intuito de apontar melhorias nos resultados. O primeiro estudo de caso realizou a combinação dos métodos através da votação média ponderada, resultando em um mapa do estado onde os municípios foram rotulados em prioritários, em transição e não prioritários, essas classes foram divididas através do valor de saída de cada localidade. O segundo estudo de caso utilizou como combinador as funções de agregação, classificando os municípios em significativos e não significativos para ocorrência do Dengue. No último estudo de caso, foram utilizadas as regras de votação baseadas em cardinalidade de conjuntos *fuzzy*, possibilitando tanto a classificação dos municípios em significativos e não significativos, quanto a construção de um mapa que dividiu os municípios em classes de graus de pertinência. Vale ressaltar, que em todos os casos houve *fuzzificação* da informação. Portanto, o resultado obtido com a aplicação da arquitetura utilizando regras de combinação *fuzzy* para métodos de aglomeração espacial, permitiu a visualização da distribuição espacial da Dengue em todos os municípios da Paraíba, direcionando os gestores para decisões que respeitem as particularidades de cada local, além de ser um método de baixo custo.

Palavras-chave: Métodos de aglomeração espacial; Epidemiologia; Conjuntos *Fuzzy*; Regra de Votação *Fuzzy*; Combinação de classificadores.

ABSTRACT

In the health area, especially in epidemiology, methods that can georeference a specific event are needed, given that the information obtained from them is essential for the understanding of the phenomenon and for the decision-making process. The combination of spatial clustering methods, with the inclusion of the uncertainties presented in the health-disease process, produces better results when compared to their individual use, in order to produce more reliable information and more precise decision-making. The objective of this study was to develop a new architecture using fuzzy combination rules for spatial clustering methods applied to epidemiology. This is an ecological, retrospective, quantitative approach. The new architecture was tested in three different case studies using epidemiological data from Dengue cases reported in the state of Paraíba, Brazil, from the years 2011 and 2017. The case studies demonstrated the usability of the proposed architecture through the grouping of methods using different rules of combination, in addition to the comparison with previous architecture in order to point improvements in the results. The first case study carried out the combination of methods through weighted majority voting, resulting in a map of the state where municipalities were labeled as a priority, transitional, and non-priority, these classes were divided by the output value for each locality. The second case study had functions of aggregation as a combinator, classifying the municipalities as significant and not significant for the occurrence of Dengue. In the last case study, voting rules based on the cardinality of fuzzy sets were used, making it possible to classify municipalities as significant or non-significant, and to construct a map that divided municipalities into classes of degrees of pertinence. It should be noted that in all cases there was information fuzzification. Therefore, the result obtained with the application of the architecture using fuzzy combination rules for spatial clustering methods allowed the visualization of the spatial distribution of Dengue in all the municipalities of Paraíba, directing the managers to decisions that respect the particularities of each place, in addition to being a low-cost method.

KEYWORDS: Spatial clustering methods; Epidemiology; Fuzzy sets; Fuzzy voting rule; Combination of classifiers.

LISTA DE FIGURAS

Figura – 1	Arquitetura geral dos métodos de aglomeração espacial	34
Figura – 2	Arquitetura de Sistema de Apoio a Decisão Espacial.....	43
Figura – 3	Arquitetura Sequencial.....	45
Figura – 4	Arquitetura Paralela.....	45
Figura – 5	Arquitetura Hierárquica.....	46
Figura – 6	Estrutura geral para combinação de classificadores.....	48
Figura – 7	Arquitetura para a combinação de métodos de aglomeração espacial...	49
Figura – 8	Nova arquitetura utilizando regras de combinação <i>fuzzy</i> para métodos de aglomeração espacial.....	60
Figura – 9	Mapa geográfico do Estado da Paraíba.....	64
Figura – 10	Percurso metodológico do estudo.....	65
Figura – 11	Aplicação da arquitetura para Combinação dos Métodos de Aglomeração Espacial utilizando o combinador Votação média ponderada para os dados do Dengue, Paraíba, Brasil.....	69
Figura – 12	Mapa da Razão de Incidência espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	70
Figura – 13	Mapa dos valores- <i>p</i> do método <i>Scan</i> espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	71
Figura – 14	Mapa dos valores- <i>p</i> do método <i>Getis Ord</i> dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	72
Figura – 15	Mapa dos valores- <i>p</i> do método <i>Besag Newell</i> dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	73
Figura – 16	Mapa de decisão da combinação dos métodos de aglomeração espacial dos casos de Dengue. Paraíba, Brasil, 2011, através da votação média ponderada.....	74
Figura – 17	Aplicação da nova arquitetura para Combinação dos Métodos de Aglomeração Espacial utilizando como combinador as funções de agregação H_{Max} , H_{Min} , H_{Med} , H_{Arit} e H_{COWA} (Funções H_{θ}) para os dados do Dengue, Paraíba, Brasil.....	76
Figura – 18	Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Min} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	77

Figura – 19	Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Max} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	78
Figura – 20	Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Med} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	79
Figura – 21	Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Arit} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	80
Figura – 22	Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{COWA} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011.....	81
Figura – 23	Mapa de decisão da combinação dos métodos de aglomeração espacial através da votação por maioria da arquitetura proposta por Holmes (2015) para o ano de 2011.....	82
Figura – 24	Aplicação da nova arquitetura para Combinação dos Métodos de Aglomeração Espacial utilizando regras de combinação baseadas em cardinalidade de conjuntos <i>fuzzy</i> para os dados do Dengue, Paraíba, Brasil.....	84
Figura – 25	Mapa da Razão de Incidência espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	85
Figura – 26	Mapa da estatística <i>Scan</i> espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	86
Figura – 27	Mapa do método Getis Ord dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	87
Figura – 28	Mapa do método Besag Newell dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	88
Figura – 29	Mapa do método Tango dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	89
Figura – 30	Mapa do método Estatística M dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	89
Figura – 31	Mapa de decisão dos métodos de aglomeração espacial utilizando a cardinalidade de conjuntos <i>fuzzy</i> . Paraíba, Brazil, 2017.....	90
Figura – 32	Mapa de decisão da votação por maioria <i>fuzzy</i> utilizando cardinalidade de conjuntos <i>fuzzy</i> dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	91

Figura – 33	Mapa de decisão da <i>Plurality voting fuzzy</i> utilizando cardinalidade de conjuntos <i>fuzzy</i> dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017.....	92
Figura – 34	Mapa de decisão da combinação dos métodos de aglomeração espacial utilizando a votação por maioria. Paraíba, Brazil, 2017.....	93
Figura – 35	Mapa de decisão da combinação dos métodos de aglomeração espacial utilizando plurality voting. Paraíba, Brazil, 2017.....	94
Figura – 36	Mapa de votação da combinação dos métodos de aglomeração espacial. Paraíba, Brazil, 2017.....	95

LISTA DE QUADROS

Quadro – 1	Principais t-normas e t-conormas.....	29
Quadro – 2	Interpretação da significância do índice local.....	37
Quadro – 3	Interpretação do RIE através das classes.....	41

LISTA DE ABREVIATURAS E SIGLAS

Sigla	Acrônimo
BN	Besag Newell
GPS	Sistema de posicionamento global
IBGE	Instituto Brasileiro de Geografia e Estatística
RIE	Razão de incidência espacial
SADE	Sistemas de apoio a decisão espacial
SIG	Sistema de Informação Geográfica
SINAN	Sistema de Informação de Agravos de Notificações
FHD	Febre hemorrágica da dengue
SCD	Síndrome do choque da dengue
DCC	Dengue com complicações
OMS	Organização Mundial da Saúde

LISTA DE NOTAÇÕES

Notação	Significado
x	Elemento de um conjunto
X	Conjunto
U	Universo
G	Conjunto de coordenadas espaciais numa região geográfica (mapa)
D	Conjuntos de dados georreferenciados
R_+	Conjunto de números reais não negativos
Ω	p -valor calculado pelo método aglomeração espacial
Ω_G	conjunto de p -valores calculados pelo método aglomeração espacial para cada geo-objeto em G
D_G	conjunto de coordenadas espacial que define uma área, com um conjunto de dados georreferenciados
$f_2 \circ f_1$	Composição de funções que definem os métodos de aglomeração espacial
L_G	Rótulo de classes para cada geo-objeto em G
z	Geo-objeto (elemento do conjunto G)
$F_{(z)}$	Razão de probabilidade da varredura Scan
H_0	Hipótese nula
H_1	Hipótese alternativa
P	Probabilidade
θ_1	Estimativa do parâmetro sob a hipótese alternativa
θ_0	Estimativa do parâmetro sob a hipótese nula
$G(d)$	Índice global
d	Distância crítica
V	matriz de proximidade da distância crítica
$v_{ij}(d)$	Elementos da matriz de proximidade da distância crítica
$G_i(d)$ e $G_i^*(d)$	Índices locais

T	Número total de casos
p	proporção esperada de casos
σ	largura da banda
M	estatística global
M_i	estatística local
w_{ij}	Matriz de pesos
M^*	Valor crítico
α	Nível de significância
z	Geo-Objeto
L	Rótulo de classes
$g_3 \circ g_2 \circ g_1$	Composição de funções da combinação de classificadores
S	probabilidade, nível de confiança ou pontuação para cada classificador
Y_s	Tipo padronizado de informação
C	Número de classificadores
Δ_{ji}	voto para a classe j registrado pelo classificador i
$\Sigma Count$	Sigma count
μ	valor de pertinência
$f_4 \circ f_3 \circ f_2 \circ f_1$	Composição de funções da combinação de métodos de aglomeração espacial.

SUMÁRIO

1	INTRODUÇÃO.....	21
1.1	JUSTIFICATIVA.....	23
1.1.1	Argumento de tese.....	24
2	OBJETIVOS.....	26
2.1	GERAL.....	26
2.2	ESPECÍFICOS.....	26
3	REFERENCIAL TEÓRICO.....	28
3.1	CONJUNTOS CLÁSSICOS E CONJUNTOS <i>FUZZY</i>	28
3.1.1	Conjuntos Clássicos.....	28
3.1.2	Conjuntos <i>Fuzzy</i>.....	28
3.2	CONCEITOS BÁSICOS DO GEOPROCESSAMENTO.....	30
3.3	EPIDEMIOLOGIA ESPACIAL.....	31
3.3.1	Epidemiologia do Dengue.....	32
3.4	MÉTODOS DE AGLOMERAÇÃO ESPACIAL.....	33
3.4.1	Estatística Scan Espacial.....	35
3.4.2	Índice de Getis Ord.....	36
3.4.3	Besag Newell.....	38
3.4.4	Estatística M.....	39
3.4.5	Método Tango.....	40
3.5	MEDIDA DE ASSOCIAÇÃO.....	40
3.5.1	Razão de Incidência Espacial.....	40
3.6	SUPORTE A DECISÃO ESPACIAL (SADE).....	41
3.7	COMBINAÇÃO DE CLASSIFICADORES.....	43
3.8	REGRAS DE VOTAÇÃO E AGREGADORES COMO MÉTODO PARA COMBINAÇÃO DE CLASSIFICADORES.....	50
3.8.1	Votação por maioria.....	50
3.8.2	Plurality Votting.....	51
3.8.3	Votação ponderada.....	51
3.8.4	Votação Flexível.....	52
3.8.5	Votação média ponderada.....	52
3.8.6	Cardinalidade de Conjuntos <i>Fuzzy</i>.....	53

3.8.7	Σ count ponderado.....	54
3.8.8	Funções H_{θ}	54
4	NOVA ARQUITETURA UTILIZANDO REGRAS DE COMBINAÇÃO FUZZY PARA MÉTODOS DE AGLOMERAÇÃO ESPACIAL APLICADA À EPIDEMIOLOGIA.....	58
4.1	NOVA PROPOSTA PARA REGRAS DE VOTAÇÃO UTILIZANDO CARDINALIDADE DE CONJUNTOS FUZZY.....	61
4.1.1	Votação por maioria fuzzy utilizando cardinalidade de conjuntos fuzzy.....	61
4.1.2	Plurality voting fuzzy utilizando cardinalidade de conjuntos fuzzy.....	61
5	PERCURSO METODOLÓGICO.....	64
5.1	ÁREA DO ESTUDO.....	64
5.2	CARACTERIZAÇÃO DO ESTUDO E ANÁLISE DOS DADOS.....	64
5.3	ASPECTOS ÉTICOS.....	66
6	RESULTADOS.....	68
6.1	ESTUDO DE CASO I.....	68
6.2	ESTUDO DE CASO II.....	75
6.3	ESTUDO DE CASO III.....	82
6.3.1	Comparação com arquitetura existente de combinação de métodos de aglomeração espacial (HOLMES, 2015).....	92
7	DISCUSSÃO.....	97
8	CONCLUSÃO.....	103
	PRODUÇÃO ACADÊMICA.....	105
	REFERÊNCIAS.....	106

*“É preciso estudar muito para
saber um pouco.”*

(Motesquieu)

APRESENTAÇÃO

A construção desta tese de doutoramento surgiu a partir das contínuas pesquisas realizadas no Laboratório de Estatística aplicada ao processamento de Imagens e Geoprocessamento (Leapig). O interesse por estudos ecológicos teve como ponto de partida os estudos de análise espacial e espaço-temporal sobre Aids, através do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) da Universidade Federal da Paraíba-UFPB. Além da construção do Trabalho de Conclusão do Curso (TCC) de Enfermagem intitulado FEMINIZAÇÃO DA AIDS NA PARAÍBA: uma análise da distribuição espaço-temporal e sócioepidemiológica no ano de 2012.

No ano de 2013 iniciei o Mestrado no Programa de Pós-Graduação em Modelos de Decisão e Saúde-UFPB, aprofundei os estudos em Análise espacial. Em 2015 com a conclusão do mestrado, construí a dissertação intitulada Um sistema de Suporte a Decisão espacial usando o perfil demográfico para o controle da Tuberculose no município de João Pessoa-PB. Este estudo apontou para a necessidade de aprofundar e compreender teoricamente a combinação de métodos de aglomeração espacial e como prestar melhores informações numa perspectiva epidemiológica.

Então, iniciei um aprofundamento na teoria de combinação de métodos de aglomeração espacial, regras de combinação e agregadores. Após embasamento teórico, foi possível a construção de uma nova arquitetura utilizando regras de combinação *Fuzzy* para métodos de aglomeração espacial aplicada à epidemiologia que apontou para obtenção de resultados satisfatórios.

Diante disso, para demonstrar a funcionabilidade da arquitetura foi utilizado o banco de dados do Dengue do estado da Paraíba, na qual apresento três estudos de casos que mostram a flexibilidade da arquitetura proposta. Sendo utilizados diferentes métodos de aglomeração espacial e regras de combinação, possibilitando uma melhor compreensão do evento estudado. Por fim, busco a construção de um aparato para a tomada de decisão espacial na área da Epidemiologia.

1 INTRODUÇÃO

A epidemiologia é definida como “a ciência que estuda o processo saúde-doença na comunidade, analisando a distribuição e os fatores determinantes das enfermidades e dos agravos à saúde coletiva, sugerindo medidas específicas de prevenção, de controle ou de erradicação”, (ROUQUAYROL, 1986, p.1). O estudo desta ciência visa prover fundamentação científica para a promoção e cuidados em saúde de acordo com as especificidades de cada localidade e de sua população, elencando os locais prioritários segundo a região geográfica na qual estão inseridos, auxiliando então a tomada de decisão em saúde (SANDERSON; GRUEN, 2006; ROTHMAN; LASH; GREENLAND, 2012).

Diante disso, ao analisar estudos sobre agravos à saúde, é possível afirmar que ao levar em consideração a distribuição espacial dos agravos a compreensão das informações é facilitada, sendo assim, a tomada de decisão dos gestores pode ser melhorada (SÁ, et al., 2018; MELO; MORAES, 2018; SÁ, 2015; MORAES, NOGUEIRA, SOUSA, 2014). Na epidemiologia, necessita-se de métodos que consigam georreferenciar o evento estudado, já que esta informação é primordial para o entendimento do fenômeno e para o processo de tomada de decisão.

Os métodos de aglomeração espacial analisam os fenômenos em estudo identificando os aglomerados espaciais significativos e não significativos. Alguns dos métodos mais utilizados na área de estatística espacial e de epidemiologia espacial são: Getis e Ord (ANSELIN, 1992), Estatística Scan Espacial (KULLDORFF; NAGARWALL, 1995), Besag e Newell (BESAG; NEWELL, 1991) Estatística M (ROGERSON, 2001) e Método Tango (TANGO, 1995). Esses métodos quando utilizados individualmente podem não refletir a realidade do fenômeno estudado, devido as diferentes metodologias em que se baseiam. Desse modo, com o intuito de obter melhores resultados a combinação de métodos de aglomeração espacial foi proposta recentemente por Holmes (2015), com base na ideia de combinação de métodos de classificação. Essa mesma concepção também está presente em outras áreas, como por exemplo, os agrupamentos em aprendizado de máquina (LANTZ, 2013).

A capacidade de generalização de um classificador é uma das propriedades mais importantes, ou seja, a capacidade de resposta ao reconhecer novos padrões (KUNCHEVA, 2004). Estudos mostram que a definição de se trabalhar com múltiplos classificadores vem com uma proposta, que visa o desenvolvimento confiável dos sistemas de reconhecimento de padrões (SUEN et al, 1990). Ao implementar classificadores distintos estes possuem diferentes limiares de decisão e generalizações. Para utilizar a informação combinada de

múltiplos classificadores, a saída de cada classificador pode ser combinada com as restantes, permitindo uma capacidade de generalização e estabilidade de classificação na decisão final (KUNCHEVA, 2004; PONTI JUNIOR, 2004).

Vários estudos utilizaram esta combinação de classificadores para problemas complexos, envolvendo conjunto de classes, conjuntos de atributos com dimensionalidades e características diferentes, já que a escolha de um único classificador causaria uma provável limitação na capacidade de reconhecimento do sistema (XU et al., 1992; KITTLER et al., 1998).

É possível encontrar nas áreas do conhecimento várias formas de combinação de classificadores descritas e distribuídas, por exemplo, reconhecimento da face, reconhecimento de pessoas, reconhecimento de caracteres, classificação de imagens, estatística espacial e entre outras (BREVE; PONTI; MASCARENHA, 2005). Do mesmo modo, são encontradas regras para combinar classificadores como, por exemplo, regras de votação (máximo, mínimo, produto, soma, votação por maioria, média e mediana) (JAIN et al., 2000).

Com a revisão da literatura foi observado que existem poucas publicações que abordam a combinação de métodos de aglomeração espacial. A primeira publicação utiliza a votação por maioria para realizar a combinação dos métodos de aglomeração realizando uma binarização da informação (HOLMES; MORAES; VIANNA, 2015). A segunda publicação generaliza o primeiro trabalho, pois desenvolve a arquitetura combinação dos métodos de aglomeração (HOLMES, 2015). O estudo mais recente aplicou a mesma arquitetura de combinação dos citados anteriormente, contudo forneceu um novo mapa flexível que considerou o número de votos recebido por cada geo-objeto (HOLMES; MORAES; VIANNA, 2018). Vale ressaltar, que uma generalização dessa arquitetura foi publicada como resultado parcial desta tese demonstrando a combinação dos métodos de aglomeração espacial através da regra de votação média ponderada, resultando em um mapa de suporte a decisão que inclui informações fuzzificadas, obtendo um ganho de informações que anteriormente não eram utilizadas (SÁ et al., 2018).

Para demonstrar a funcionabilidade da arquitetura proposta foram utilizados os dados do Dengue no estado da Paraíba, Brasil. O dengue é uma doença viral sistêmica e dinâmica e tem um amplo espectro clínico que inclui tanto as manifestações clínicas graves e não-graves, transmitida por mosquito Aedes e sua incidência tem aumentado cerca de 30 vezes ao longo dos últimos 50 anos (WHO, 2015).

Atualmente este agravo, representa um problema de saúde pública para cerca da metade da população mundial, aproximadamente 3,9 bilhões de pessoas em 128 países, sendo

considerada, uma doença complexa, e faz-se necessária uma compreensão mais profunda das interações entre vírus, hospedeiro e vetor, e os custos da doença na saúde pública. A dengue pode afetar qualquer pessoa independentemente de idade, sexo, condição de saúde ou status socioeconômico. A distribuição geográfica dos quatro sorotipos da dengue varia imprevisivelmente ao longo do tempo, provocando risco contínuo de infecção em populações de áreas endêmicas, apesar das ações de controle de vetor (WHO, 2016).

A utilização dos métodos de aglomeração espacial na detecção de aglomerados significativos para um determinado fenômeno epidemiológico é fundamental para direcionar políticas e ações segundo a particularidade de cada localidade. Nesse contexto, a combinação desses métodos aponta para avanços, pois seus resultados aproximam-se da realidade do fenômeno estudado ao preservar e utilizar o máximo de informação possível tornando o suporte a decisão mais preciso. Diante do exposto, este trabalho propõe uma inovação nas áreas de estatística espacial e epidemiologia que é o desenvolvimento de uma nova arquitetura utilizando regras de combinação fuzzy para métodos de aglomeração espacial aplicada à epidemiologia

1.1 JUSTIFICATIVA

Do ponto de vista epidemiológico, a escolha do Dengue para aplicação da arquitetura justifica-se pela grande necessidade de detecção de aglomerados espaciais da doença na região geográfica, possibilitando um combate à doença mais efetivo devido a detecção da presença do vetor.

O estudo realizado por Holmes (2015), traz como resultado a construção de uma arquitetura que realiza a combinação de métodos de aglomeração espacial aplicado a epidemiologia, apontando vantagens da combinação em relação a utilização de um método de aglomeração isoladamente, fornecendo também resultados interessantes para a epidemiologia do dengue na Paraíba (HOLMES;MORAES;VIANNA,2015; HOLMES, 2015; HOLMES;MORAES;VIANNA, 2018). Ao analisar estes estudos encontra-se uma limitação na arquitetura que foi a binarização da informação sobre p-valores significativos e não significativos resultantes de métodos de aglomeração espacial, pois leva a uma perda de informações, comparado quando utilizados dados fuzzificados.

Diante disso, após uma vasta busca na literatura não foram encontrados estudos que abordam a combinação de métodos de aglomeração espacial que utilizem regras de combinação *fuzzy*. Sendo assim, ao propor uma nova arquitetura utilizando regras de

combinação *fuzzy* para métodos de aglomeração espacial aplicada à epidemiologia, tornou-se possível a utilização de toda informação, potencializando o conhecimento do processo de saúde-doença, pois adiciona as incertezas ligadas a esse processo.

O funcionamento dessa nova arquitetura foi testado através de estudos de casos que utilizaram diferentes métodos de aglomeração e regras de combinações, permitindo o uso da informação em sua totalidade, através dos graus de pertinência ou a utilização da informação binária quando necessário.

1.1.1 Argumento de Tese

Defende-se a tese que a combinação dos métodos de aglomeração espacial utilizando regras de combinação *fuzzy*, permite a inclusão das incertezas presentes no processo saúde-doença, preservando assim informações, possibilitando resultados mais fidedignos e tomada de decisões mais precisas.

“Se a educação sozinha não transforma a sociedade e, sem ela tampouco a sociedade muda”.

Paulo Freire

2 OBJETIVOS

2.1 GERAL

Desenvolver uma nova arquitetura utilizando regras de combinação *fuzzy* para métodos de aglomeração espacial aplicada à epidemiologia

2.2 ESPECÍFICOS

- Propor uma nova arquitetura para combinação de métodos de aglomeração espacial;
- Formalizar matematicamente a nova arquitetura para combinação de métodos de aglomeração espacial;
- Propor novas regras de combinação *fuzzy* para métodos de aglomeração espacial;
- Analisar os dados epidemiológicos do Dengue espacialmente, aplicando diferentes métodos de aglomeração espacial;
- Demonstrar o funcionamento da nova arquitetura para combinação de métodos de aglomeração espacial, aplicando aos dados epidemiológicos do Dengue da Paraíba.

“O educador se eterniza em cada ser que educa”.

Paulo Freire

3 REFERENCIAL TEÓRICO

Este capítulo, contém o aporte teórico para a construção da nova arquitetura de combinação. Para isso, permeará os conceitos básicos de conjuntos, geoprocessamento, aprofundando nos métodos de aglomeração espacial mais utilizados, epidemiologia espacial, Sistemas de apoio a decisão espacial (SADE). Em seguida, serão abordadas a combinação de classificadores e as regras de combinação e agregadores.

3.1 CONJUNTOS CLÁSSICOS E CONJUNTOS *FUZZY*

3.1.1 Conjuntos Clássicos e Lógica de primeira ordem

As ideias essenciais da teoria dos conjuntos foram introduzidas por G. Cantor no ano de 1895. Define-se conjunto como uma coleção de objetos de natureza qualquer, dentro de um todo de objetos definidos e distinguíveis, os quais se dizem elementos do conjunto. Representa-se por $x \in X$ a condição “ x é um elemento do conjunto X ” que também se lê “ x pertence a X ”. O complemento desta condição escreve-se $x \notin X$.

A teoria de Tipos desenvolvida por Bertrand Russell dividia os objetos do discurso matemático em níveis, destacando-se: o nível zero (ou ordem zero), que é denotado como Cálculo Proposicional, e o nível 1 (ou primeira ordem) que é a Lógica de Primeira Ordem. Então uma linguagem de primeira ordem consiste num alfabeto que contém os símbolos lógicos $\wedge$ (e), $\vee$ (ou), $\neg$ (não), $\rightarrow$ (implica), $\leftrightarrow$ (equivale), $\exists$ (existe) e $\forall$ (qualquer), e também o da igualdade, “=”; um conjunto enumerável de símbolos de variáveis; além disso a linguagem tem regras (gramaticais) de formação de expressões, ou fórmulas e sentenças. Então, se A e B são fórmulas, temos as seguintes correspondências formais para as proposições compostas: $\neg A$ (negação de A), $A \wedge B$ (conjunção de A e B), $A \vee B$ (disjunção de A e B) e $A \rightarrow B$ (condicional de A para B) (CASTRUCCI, 1975; BITTENCOURT, 1996).

3.1.2 Conjuntos *Fuzzy*

Define-se um conjunto *Fuzzy* como uma classe de objetos com contínuos graus de associação. Esse conjunto caracteriza-se por funções de pertinência, atribuídas a cada objeto do conjunto, as quais variam entre zero e um. Logo, sendo X um conjunto no universo U e x um elemento pertencente a U , a função de pertinência de um conjunto *Fuzzy*, associa a cada

ponto em U , um número real presente no intervalo de 0 a 1, ou seja, $X:U \rightarrow [0,1]$ (ZADEH, 1965).

Nos conjuntos *Fuzzy*, é possível realizar operações de intersecção e união entre os conjuntos. Para tal, são utilizados os operadores t-norma e t-conorma. O primeiro operador, t-norma, permite a intersecção entre os conjuntos *Fuzzy*, sendo representado pelo conector “E”, enquanto a t-conorma permite a união entre os conjuntos *Fuzzy*, representado pelo conector “OU” (MORAES; BANON; SANDRI, 2002). As principais t-normas e t-conormas estão expressas no quadro 1.

Quadro 1 - Principais t-normas e t-conormas

T-normas	T-conormas
$\min(a, b)$	$\max(a, b)$
$a \times b$	$a + b - ab$
$\max(a + b - 1)$	$\min(a + b, 1)$
a , se $b = 1$ b , se $a = 1$ 0 , senão	a , se $b = 0$ b , se $a = 0$ 1 , senão

Fonte: Adaptação Moraes; Banon; Sandri, 2002

Uma das formas mais comuns de expressar este conhecimento é por meio de regras do tipo condição-ação. Dessa forma, todo o conhecimento é representado por meio de um conjunto de regras onde as condições são dadas a partir de um conjunto de termos linguísticos, os quais são associados às variáveis de entrada/saída. Pode-se entender a inferência baseada em regras *Fuzzy* como um funcional caracterizado por mapear um conjunto de entradas do sistema para um conjunto de saídas. Portanto, a regra *Fuzzy* constitui uma unidade capaz de capturar um conhecimento específico (PASSINO, 1998).

Essas regras possuem uma estrutura básica definida em duas partes: O Antecedente (SE), que é composta pelas variáveis de entrada que descrevem uma condição, e o Consequente (ENTÃO), composto pelas variáveis de saída, indicando uma conclusão. Vale salientar que no Antecedente é permitido o uso de mais de um conector como o “E” e “OU” e

no Consequente pode haver mais de uma conclusão (ORTEGA, 2001). De forma genérica, uma regra Fuzzy pode ser representada da seguinte forma:

$$\text{SE } (x \text{ é } a_i) \text{ E } (y \text{ é } b_i) \text{ OU... ENTÃO } (r \text{ é } c_i) \text{ E } (w \text{ é } d_i)...$$

onde x e y são as variáveis linguísticas de entrada, r e w as de saída, e a_i , b_i , c_i e d_i são realizações dessas variáveis, medidas a partir da interação do especialista com o sistema (GOMIDE; GUDWIN, 1994).

Sadegh-Zadeh propõe que as teorias e conceitos que permeiam a saúde e o adoecimento, não são passíveis da classificação pela lógica clássica, então ele sugere uma teoria difusa, utilizando uma lógica não-clássica, por meio do método *fuzzy* (ZADEH, 2000).

3.2 CONCEITOS BÁSICOS DO GEOPROCESSAMENTO

Geoprocessamento “é o conjunto de técnicas de coleta, tratamento e exibição de informações referenciadas geograficamente, funcionando como uma ferramenta de visualização de eventos epidemiológicos em mapas.” (SANTOS, et al, 2004). Destacam-se: sensoriamento remoto, digitalização dos dados, automação de tarefas cartográficas, Sistema de Posicionamento Global (GPS) e Sistema de Informação Geográfica (SIG) (SANTOS, et al, 2004).

O SIG tem como objetivo principal, representar em ambiente computacional os fenômenos geográficos, que se pretende estudar (CÂMARA et al., 2001). A aplicação do SIG na pesquisa em saúde oferece grandes possibilidades, oferecendo aos pesquisadores novos métodos para o manejo de sua informação espacial, tornando-se uma poderosa ferramenta para conexão entre saúde e ambiente (SANTOS, et al, 2004).

Para isso é necessário compreender conceitos inerentes a área, sendo um desses conceitos o **espaço geográfico** que pode ser definido, como sendo uma coleção de localizações na superfície da Terra, onde ocorrem os fenômenos geográficos, portanto definindo-se, em função de suas coordenadas, sua altitude e sua posição relativa, possível de ser cartografado (DOLFUD, 1991).

O espaço geográfico é modelado no geoprocessamento segundo duas visões complementares: os geo-campos e os geo-objetos (WORBOYS, 1995).

- **Geo-campo** representa a distribuição espacial de uma variável, que possui valores em todos os pontos pertencentes a uma região do espaço geográfico, num dado tempo t (CÂMARA et al., 2001).
- **Geo-objetos** (ou **objetos geográficos**) são entidades distintas e localizáveis que compõem uma determinada região do espaço geográfico, ou seja, são os componentes urbanos identificáveis (WORBOYS, 1995). Nessa tese, os geo-objetos são os municípios do Estado da Paraíba

3.3 EPIDEMIOLOGIA ESPACIAL

A pesquisa de dados espaciais em saúde iniciou-se no século XIX, com John Snow, que buscou analisar a epidemia da cólera na região de Soho em Londres, em 1854, conseguindo encontrar uma associação dos mortos por cólera e as bombas públicas que abasteciam a cidade, apontando a causa da epidemia (MEDRONHO et al., 2009).

Na área da epidemiologia percebeu-se a necessidade de formulação de estratégias de controle de agravos, então várias metodologias têm sido utilizadas na caracterização epidemiológica das doenças. A utilização de metodologias que melhor destaca processos ambientais e sociais que interferem nos padrões de transmissão de doenças é de grande importância para adoção de medidas eficazes de prevenção e controle. Assim, a estratificação do espaço, segundo indicadores socioambientais, acrescida das informações relativas ao nível de endemidade da área, constitui um instrumento de apoio ao planejamento das ações de controle (RESENDES, 2010).

Os eventos relacionados a saúde, tais como doenças, mortes, nascimentos, risco de exposição ou outro evento, podem ser referenciados geograficamente. O estudo da distribuição geográfica e associação espacial dos eventos de saúde pode ser chamado de epidemiologia espacial, que na área da saúde traz importantes contribuições para a compreensão dos processos de saúde-doença e pode levar a resultados diferentes daqueles obtidos em estudos em que não é considerado o espaço geográfico. Então, além da descrição das características das populações é necessário identificar onde ocorrem os casos, localizando o risco potencial das áreas ambientais e sociais vulneráveis (CARVALHO; SOUZA; SANTOS, 2005).

Sendo assim, as técnicas estatísticas tradicionais usadas para analisar dados sociais, econômicos, ambientais e de saúde, embora sejam importantes e úteis não associam as

ocorrências num contexto espacial. Como o processo saúde-doença é um resultado direto dos fatores de risco biológicos, sociais e ambientais que influenciam a saúde e bem-estar das populações precisam ser integradas dentro de estudos que utilizam a epidemiologia espacial onde os indivíduos são vistos em seu ambiente social, econômico e cultural, juntamente com a carga biológica (SUSSER, 1996). O uso dos métodos estatísticos espaciais pode auxiliar as autoridades públicas numa avaliação mais rápida, confiável e válida em saúde. Tudo isso contribui para um melhor entendimento da compreensão dos problemas de saúde da atualidade (MEDRONHO et al., 2009).

A aplicação da estatística espacial e SIG para os resultados de saúde já está sendo usado para proporcionar novas formas de examinar os padrões geográficos da doença. Na epidemiologia espacial, o espaço geográfico é essencial para entender a propagação e distribuição dos agravos, tendo a finalidade de descrever, quantificar e explicar as variações geográficas da doença, avaliar a associação entre a incidência e os fatores de risco e identificar o potencial de agregação especial das enfermidades (ELLIOT; WAKEFIELD, 2000). Levando em consideração o fato de que praticamente todas as doenças podem variar com a área geográfica. Essa variação deve ser analisada sistematicamente, a fim de entender se esta é condicionada por fatores ambientais ou sociais (BHOPAL, 2007).

Na área da epidemiologia espacial é possível encontrar aplicações em diversas áreas e utilizando métodos variados, como exemplos tem-se os estudos que analisaram a Epidemiologia espacial de violência entre parceiros íntimos (GRACIA et al., 2015), Análise espacial do Dengue (HOLMES, 2015; SÁ et al., 2018), Epidemiologia espacial de picadas de escorpiões (ARAÚJO et al., 2017), entre outros.

3.3.1 Epidemiologia do Dengue

O Dengue é um tipo de arbovirose, ou seja, doença transmitida por um mosquito, neste caso o *Aedes aegypti*, causada por quatro sorotipos de vírus do dengue (DENV-1, DENV-2, DENV-3 e DENV- 4) (WHO, 2016). O mosquito *Aedes aegypti*, que é o vetor transmissor do dengue, foi detectado nas principais cidades do Brasil na década de 1970, depois de ter sido erradicado na década de 1950. Até 2013, a classificação dos casos no Brasil se dividia em febre hemorrágica da dengue (FHD), síndrome do choque do dengue (SCD) e dengue com complicações (DCC). A partir de 2014, começou a adotar a nova classificação de casos do dengue da Organização Mundial da Saúde (OMS), sendo

atualmente classificados como dengue com sinais de alarme e dengue grave (BRASIL, 2014).

Estimativas da OMS apontam que em 2014, cerca de 2,3 milhões de pessoas tiveram Dengue no mundo. A taxa de incidência registrada ficou em torno de 455,4 por 100.000 habitantes (hab.) (WHO, 2015). No Brasil, foram notificados 1.452.489 casos novos em 2013 no Sistema de Informação de Agravos de Notificação (Sinan). Em relação ao coeficiente de incidência (CI) o ano de 2011 apresentou uma taxa 400/100.000 hab., com uma diminuição em 2012 com o valor de 303,9/100.000 hab. e crescimento relevante em 2013 com uma taxa de 722,4/100.000 hab (BRASIL, 2014).

Na Paraíba em 2014, foram notificados 7.366 casos suspeitos de dengue, destes 1.831 foram descartados e 3.442 foram confirmados. Em comparação ao número de notificações de 2013, com 18.078 notificações, percebe-se uma redução de 59,26%. Levando em consideração o cenário nacional, a Paraíba encontra-se em conformidade, visto que, o número de casos registrados de dengue no Brasil caiu 61% entre janeiro e 15 de novembro de 2014, em comparação ao mesmo período de 2013. Foram 1,4 milhão de casos em 2013 contra 566,6 mil neste ano, no mesmo período. Todas as regiões do país apresentaram redução de casos notificados, sendo que a região Sudeste teve a queda mais representativa, correspondente a 67%, seguida pelo Sul (64%), Centro-Oeste (58%), Nordeste (42%) e Norte (12%) (BRASIL, 2014). Embora esteja acontecendo uma redução, o número de casos ainda é bem representativo e os fatores que contribuem para este grande quantitativo de casos são as áreas urbanas empobrecidas, subúrbios e zona rural, mas também afeta bairros mais ricos em países tropicais e subtropicais (WHO, 2015).

Nesse contexto, com a grande incidência de casos de Dengue no Estado da Paraíba, optou-se por realizar os estudos de casos referentes a esta morbidade para entender como essa se distribui espacialmente, já que o combate à doença é mais efetivo se for possível detectar a presença do vetor. Do ponto de vista epidemiológico, isso se concretiza a partir da detecção de aglomerados espaciais da doença na região geográfica em estudo e promovendo medidas de combate direcionadas aquela localidade.

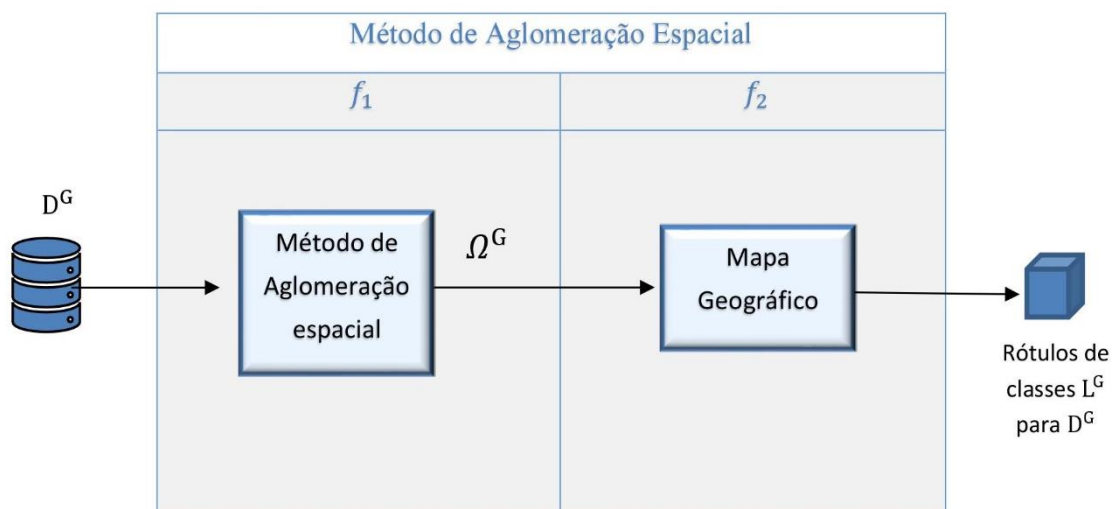
3.4 MÉTODOS DE AGLOMERAÇÃO ESPACIAL

Na epidemiologia, as pesquisas que utilizam métodos de análise espacial têm avançado, visto que analisar as doenças segundo a sua localização no espaço e no tempo relaciona o homem e o processo saúde-doença com o ambiente no qual ele vive. Essas

pesquisas têm sido descritas devido à sua importância para a saúde pública pela sua capacidade de mapear a distribuição da doença e fatores de risco a níveis populacionais (SOUZA et al., 2007). Nesse contexto, a utilização dos métodos aglomeração espacial se mostra como uma ferramenta muito útil para o estudo das dinâmicas da distribuição espacial que permite detectar as regiões geográficas que apresentam maior risco epidemiológico (SKALINSKI; COSTA; TEIXEIRA, 2019).

Um método de aglomeração espacial, pode ser visto como uma composição de funções $f_2 \circ f_1$, como ilustrado na Figura 1. Formalmente, seja $G \in R_+^2$, onde R_+ (o conjunto de números reais não negativos), seja um conjunto de coordenadas geográficas numa região, sobre a qual os geo-objetos z são apresentados, $G = \{z_1, \dots, z_k\}$, ou seja, é um mapa geográfico. Considerando, D um conjunto de dados georreferenciados relacionado a um geo-objeto em G e $\Omega \in [0,1]$ é o p -valor calculado pelo método aglomeração espacial para o geo-objeto. Tem-se que, o domínio da primeira função da composição, denotada por f_1 , é o conjunto de coordenadas espaciais que define uma área, com um conjunto de dados georreferenciados, que é denotado por D^G . O co-domínio para f_1 é o conjunto de p -valores calculados pelo método aglomeração espacial para cada geo-objeto em G , que é denotado por Ω^G (Figura 1). O mapa final é fornecido pela segunda função da composição, denotada por f_2 , que tem seu domínio em Ω^G e seu co-domínio é um conjunto de rótulos de classe L^G (Figura 1).

Figura 1 - Arquitetura Geral dos Métodos de Aglomeração Espacial



Fonte: Elaborada pelo próprio autor, 2019

Foram utilizados nesta tese alguns dos métodos mais utilizados na área de estatística espacial e de epidemiologia espacial, sendo descritos detalhadamente a seguir. Estes são: Getis e Ord (ANSELIN, 1992), Estatística Scan Espacial (KULLDORFF; NAGARWALL, 1995), Besag e Newell (BESAG; NEWELL, 1991) Estatística M (ROGERSON, 2001) e Método Tango (TANGO, 1995).

3.4.1 Estatística Scan Espacial

O Scan circular é um caso da estatística Scan espacial, sendo atualmente a ferramenta mais utilizada para a detecção e inferência de aglomerados espaciais de doenças. O modelo probabilista de Poisson é geralmente o mais utilizado em modelagem epidemiológica. Em sua metodologia, é gerada uma janela na qual associa-se a informação da área a um único ponto dentro do polígono, o centróide, que é o centro de massa de cada geo-objeto. A seguir, procuram-se os geo-objetos onde a ocorrência do evento seja significativamente mais provável de ocorrer dentro de uma área do que fora dela. Neste caso, uma janela consiste no círculo criado a partir do centroide, que varia continuamente desde zero até um percentual da população em risco a ser coberta (KULLDORFF; NAGARWALLA, 1995).

A varredura Scan é baseada na razão de probabilidade $F(z)$ com as seguintes hipóteses: H_0 consiste em que não há conglomerados espaciais na região de análise e a $H_1(z)$, onde o geo-objeto z detecta conglomerado espacial. Estas hipóteses são baseadas na razão de verossimilhança dada por $F(z)$ para um círculo no geo-objeto z , sendo definida entre a razão de verossimilhança das informações sob a hipótese alternativa $H_1(z)$, seguido pela verossimilhança sob a hipótese nula H_0 (NEILL; MOORE; SABHNANI, 2005; KULLDORFF; NAGARWALLA, 2006):

$$F(z) = \frac{P(\text{Dados}|H_1(z))}{P(\text{Dados}|H_0)} \quad (1)$$

Caso a H_0 ou H_1 tenha parâmetros livres, a razão de verossimilhança é dada por:

$$F(z) = \frac{\max_{\theta_1(z) \in \theta_1(z)} P(\text{Dados}|H_1(G), \theta(G))}{\max_{\theta_0 \in \theta_0} P(\text{Dados}|H_0, \theta_0)} \quad (2)$$

onde $G = \{z_1, \dots, z_k\}$ é o conjunto de todos os possíveis candidatos a conglomerados; θ_1 é uma estimativa do parâmetro para o círculo z pertencente ao conjunto de todas as estimativas pertencentes dos parâmetros $\theta_1(z)$; θ_0 é a estimativa do parâmetro sob a hipótese nula e θ_0 é o conjunto de todas as estimativas sob a H_0 .

Uma vez encontradas as regiões com os maiores escores de $F(z)$, é necessário determinar a significância estatística destas regiões. A significância estatística do conglomerado mais provável dos casos observados é calculada por meio das simulações de Monte Carlo com $\alpha = 0,05$ (LUCENA; MORAES, 2009; MOURA, 2006).

No método Scan a restrição é dada em relação ao percentual da população sob risco. De modo que, determinamos que o raio de busca seja aumentado até que contenha no máximo $\beta\%$ da população. Não há padronização com relação a este percentual da população, procura-se aquele que melhor se adeque a cada situação estudada (LUCENA; MORAES, 2012). Nesse estudo, optou-se pela comparação do mapa proveniente do Scan com o mapa RIE, onde o percentual da população foi o que forneceu resultados próximos da distribuição espacial do mapa RIE.

3.4.2 Índices de Getis e Ord

Getis e Ord introduziram as estatísticas G_i e G_i^* como meios alternativos de medidas de associação espacial, estes podem ser aplicados a dados onde a distribuição é não-normal, tendo por objetivo fazer uma medida de autocorrelação espacial não-paramétrica. As estatísticas de Getis Ord são estimadas por grupos de vizinhos da distância crítica d de cada área i . A distância crítica formada por uma matriz de proximidade V , cujos elementos são formados em função da distância crítica $v_{ij}(d)$.

Eles propuseram duas funções estatísticas: o índice global $G(d)$, que é similar às medidas tradicionais de autocorrelação espacial, e os índices locais $G_i(d)$ e $G_i^*(d)$, que são medidas de associação espacial para cada área i (ANSELIN, 1992).

Nesse caso, utilizaremos os índices locais $G_i(d)$ e $G_i^*(d)$ que são uma medida de associação para cada unidade espacial individual para cada observação i , onde estes índices indicam a extensão a que essa posição é cercada por valores elevados ou por valores baixos para a variável (ANSELIN, 1992).

A estatística local mede a associação em cada unidade espacial para cada observação i , em que $G_i(d)$ e $G_i^*(d)$ mostram a posição cercada por valores altos ou baixos para a variável. A equação do $G_i(d)$ para cada observação i e distância d é mostrada da seguinte forma:

$$G_i(d) = \frac{\sum_j v_{ij}(d)x_j}{\sum_j x_j} \quad (3)$$

onde a soma é realizada sobre todas as posições de j excluindo-se a posição i .

A equação do $G_i^*(d)$ para cada observação i e distância d é mostrada da seguinte forma:

$$G_i^*(d) = \frac{\sum_j v_{ij}(d)x_j}{\sum_j x_j} \quad (4)$$

onde a soma é realizada sobre todas as posições, inclusive a posição i .

A seguir está descrito a interpretação da significância do índice local (Quadro 2).

Quadro 2 - Interpretação da significância do índice local

Negativos***	Índice Negativo com p-valor menor que 0,005;
Negativos**	Índice Negativo com p-valor entre 0,005 e 0,025;
Negativos*	Índice Negativo com p-valor entre 0,025 e 0,05;
Negativos	Índice Negativo com p-valor acima de 0,05;
Positivos	Índice Positivo com p-valor acima de 0,05;
Positivos*	Índice Positivo com p-valor entre 0,025 e 0,05;
Positivos**	Índice Positivo com p-valor entre 0,005 e 0,025;
Positivos***	Índice Positivo com p-valor menor que 0,005;

Fonte: Elaborado pelo autor, 2017

Portanto, o índice local é interpretado da seguinte forma: os valores padronizados positivos e significantes (valor-p menor que 5%) informam uma aglomeração espacial de valores altos. Os valores das estatísticas padronizados negativos e significantes (valor-p menor que 5%) indicam uma aglomeração espacial de valores baixos.

3.4.3 Besag Newell

Besag e Newell propuseram um método com uma saída visual de conglomerados verossímeis de formato circular que tomam por base círculos centrados sobre os centróides de cada geo-objeto, com a possibilidade de configurar vários parâmetros (OPENSHAW et al., 1988). Denotado por BN, o método determina o raio necessário para que o círculo contenha pelo menos t de casos em seu interior. O procedimento é iniciado com um círculo de raio igual a zero. Se este contém t ou mais casos, o processo é interrompido; caso contrário, o raio é aumentado até incluir o centróide mais próximo. Logo, são adicionados os respectivos casos e a população exposta ao risco. Deste modo, os raios são definidos de tal forma a incluir novos centroides quando necessário, sendo executado este procedimento até que haja pelo menos t casos inseridos no círculo (BESAG; NEWELL, 1991).

Seja T o número total de casos em toda região de estudo e R a população total exposta ao risco na região. Sejam ainda $T_{j(i)}$ e $R_{j(i)}$ o número de casos e a população acumulada das j áreas mais próximas ao centróide i . A estatística do teste baseia-se na variável aleatória Q , definida como o mínimo de áreas próximas necessárias para que se obtenha um número t de casos mais próximos do centróide. Com isso, temos (BESAG, NEWELL, 1991):

$$Q = \min \{j: T_{j(i)} \geq t\} \quad (5)$$

Um aglomerado é significativo quando o valor- p , $\Omega_{t(i)}$, é menor que a significância adotada. Portanto, a partir do valor b_i , observado para Q_i , o nível de significância do teste é definida por $P(Q_i \leq q_i) = 1 - P(Q_i > q_i)$, que busca a hipótese nula, que é a hipótese de ausência de cluster espacial. O $\Omega_{t(i)}$ é definido por (BESAG; NEWELL, 1991):

$$\Omega_{t(i)} = P(Q_i \leq q_i) = 1 - \sum_{j=1}^{t-1} \frac{(R_{j(i)}T|R)^j}{j!} \exp(-R_{j(i)}T|R) \quad (6)$$

onde R_j é a população observada em l áreas.

A interpretação é feita da seguinte forma, se o p -valor obtido é menor que a significância adotada o aglomerado é dito significativo. Assim, após a obtenção de todos os

círculos que englobam um número t casos por meio do método BN, são desenhados no mapa apenas os círculos significativos ($p\text{-valor} < \alpha$). Geralmente a escolha de α é feita para permitir os muitos testes simultâneos. São considerados valores menores que a significância usuais 0,05 ou 0,01 (COSTA; ASSUNÇÃO, 2005).

3.4.4 Estatística M

A *Estatística M* é um método de aglomeração espacial no qual é aplicada a estatística global e posteriormente, dependendo do resultado obtido, a local. Segundo Lee, Yamada e Rogerson (2007) a estatística global M e a local M_i são dadas por:

$$M = \max_i \left(M_i = \frac{\sum_{j=1}^n w_{ij} y_j}{\sqrt{\sum_{j=1}^n w_{ij}^2}} \right), \quad (7)$$

onde y_j é a variável normal na região j ; a estatística M_i local é o *y-score* ponderado espacialmente em uma localidade i ; obs_j são os valores observados e exp_j são os valores esperados de cada sub-região e w_{ij} é a matriz gaussiana de pesos entre as regiões i e j .

A detecção de aglomerados espaciais é dada pelo valor crítico (M^*):

$$M^* = \sqrt{-\sqrt{\pi} \ln \left(\frac{4\alpha(1 + 0,81\sigma^2)}{n} \right)}, \quad (8)$$

onde α é o nível de significância; σ é a largura de banda, cuja escolha é arbitrária e n é o número de sub-regiões. Quando a *Estatística M* for menor ou igual ao valor crítico (M^*), não se rejeita a hipótese nula de que não há conglomerado espacial.

Para este estudo foi utilizado o Método de Poisson, pois segundo Rogerson (2001), ele fornece, na maioria das vezes, os melhores resultados, pois sua equação é consistente com os componentes da equação global do teste.

3.4.5 Método Tango

Este método foi proposto por Tango (1995) inicialmente. Sendo que o método analisava a quantidade de observações de cada sub-regiões, não analisando a diferença entre o número de observações e o número esperado de observações, levando em consideração a população de cada sub-regiões. Neste método os dados não precisam possuir distribuição normal.

Neste caso foram calculados os aglomerados de cada geo-objeto da Paraíba. O método proposto por Tango é dado pela seguinte estatística:

$$Q_S = (T - p)' B (T - p) \quad (9)$$

em que T são os vetores observados, em que $T_i = t_i/t$ e $t = t_1 + t_2 + \dots, t_k$, é o número de casos por regiões, p é a proporção esperada de casos, B é uma matriz de $n \times n$ elementos (ROGERSON et al, 1997).

A estatística de Tango é dada pela aproximação da distribuição normal e pela distribuição qui-quadrado. Neste trabalho abordaremos a aproximação pela distribuição qui-quadrado, que utiliza valores baixos de σ . Pois, a medida que a largura da banda (σ) se aproxima de zero, o valor da estatística de Tango também se aproxima do uso da distribuição qui-quadrado (ROGERSON et al, 1997).

3.5 MEDIDA DE ASSOCIAÇÃO

3.5.1 Razão de incidência espacial

A razão de incidência espacial (RIE) é um indicador que representa a ocorrência de um fenômeno de um geo-objeto com relação a toda região de estudo. Formalmente, seja uma região geográfica "G" formada por uma coleção de geo-objetos denotados por $z_1, z_2, \dots, z_n$. Seja ainda Y , uma variável aleatória que corresponde à contagem de ocorrências de um evento epidemiológico dentro de cada geo-objeto em G, denotadas por $y_1, y_2, \dots, y_n$. Por fim, seja ρ a população sob risco para aquele evento epidemiológico em cada geo-objeto de G, denotadas por $m_1, m_2, \dots, m_n$.

Então, a RIE para um geo-objeto z_i é dada por:

$$RIE(z_i) = \frac{y_i/m_i}{\sum_{j=1}^n y_j / \sum_{j=1}^n m_j} \quad (10)$$

ou seja, a RIE é dada pela incidência de um evento epidemiológico no geo-objeto z_i dividido pela incidência em toda a região geográfica.

Logo após, faz-se a divisão do resultado obtido em classes (quadro 3), e com isso, gera-se um mapa coroplético que associa uma cor a cada intervalo pré-estabelecido. Tais mapas permitem comparar informações de diferentes áreas, pois padroniza os dados retirando, assim, o efeito das diferentes populações. Este indicador representa a intensidade da ocorrência de um fenômeno com relação a todas as regiões de estudo (FREITAS; HOLMES; MORAES, 2012; LUCENA; MORAES, 2012; MORAES; NOGUEIRA; SOUSA, 2014; HOLMES; MORAES; VIANNA, 2015).

Quadro 3 - Interpretação do RIE através das classes

RIE	Interpretação do RIE
$0 \leq x < 0,5$	Risco muito baixo
$0,5 \leq x < 1$	Risco baixo
$1 \leq x < 1,5$	Risco médio
$1,5 \leq x < 2$	Risco alto
$x \geq 2$	Risco muito alto

Fonte: Quadro adaptado (COSTA, 2011)

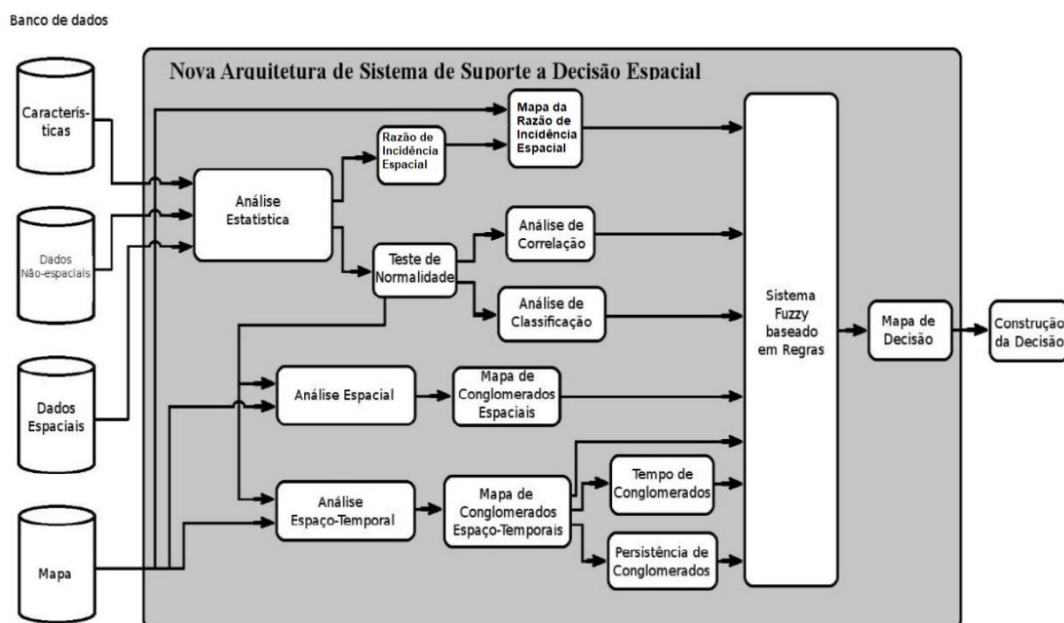
3.6 SISTEMA DE SUPORTE A DECISÃO ESPACIAL (SADE)

Embora não exista um consenso sobre uma definição para o processo de tomada de decisão espacial, é possível dizer que é um processo de tomada de decisão em que a informação espacial deve ser tomada em conta para a decisão final (SUGUMARAN, 2011). Da mesma forma o SADE não tem uma definição universalmente aceita, mas pode ser entendida como uma tecnologia computacional que fornece suporte para a tomada de decisões sobre os problemas em que há um componente geográfico ou espacial que afeta a decisão final (MORA; FORGIONE; GUPTA, 2002).

Uma outra definição para os SADE pode ser descrita como sistemas integrados de computador que auxiliam os tomadores de decisão a tratar problemas espaciais semiestruturados ou não estruturados de forma interativa e não interativa com funcionalidade para lidar com bancos de dados espaciais e não espaciais, recursos de modelagem analítica, suporte à decisão, utilitários como análise de cenário e utilitários eficazes de apresentação de dados e informações (SUGUMARAN; DEGROOTE, 2011). Esta metodologia vem sendo utilizada para auxiliar nas decisões pois é um recurso que combina características espaciais e não espaciais no processo de tomada de decisão, tornando-se essenciais no processo de tomada de decisão final (KEENAN; JANKOWSKI, 2018).

Os SADE são comumente aplicados em múltiplos cenários (MELO; MORAES, 2018). Na epidemiologia foi desenvolvido uma arquitetura de SADE por Moraes, Nogueira e Sousa (2014), que utiliza uma arquitetura levando em consideração aspectos epidemiológicos para tomada de decisão na gestão de saúde pública usando dados representados por elementos de área, isto é, quando a posição geográfica é desconhecida, mas pode ser determinado o valor total de ocorrências de cada área. Nesta arquitetura o banco de dados é analisado e dividido em categorias que podem ser: características, dados espaciais, dados não espaciais e mapas. Então são analisados segundo RIE, varredura SCAN espacial, SCAN espaço-temporal, Correlação de Spearman, método de classificação, servindo de entrada para um sistema baseado em regras *Fuzzy*. Os resultados fornecidos por este SADE são mapas de decisão em que são apontadas subáreas prioritárias, com diferentes níveis de prioridade. Esta arquitetura foi aplicada a primeira vez em sua totalidade em estudo realizado por Sá, Nogueira e Moraes (2015), onde foi realizado uma análise da Tuberculose em João Pessoa-PB (Figura 2).

Figura 2 – Arquitetura de Sistema de Apoio a Decisão Espacial



Fonte: SÁ, 2015.

Uma generalização da arquitetura foi proposta por Soares (2018), para criação de um modelo que possibilitasse a definição de níveis de prioridade de intervenção municipal para redução da mortalidade infantil no estado da Paraíba, acrescentando a Taxa de Mortalidade Infantil e agregou-se mais uma saída, o mapa de conglomerados focado no tempo de sobrevivência das crianças. Esta arquitetura SADE caracterizou-se por apresentar uma saída simples de fácil interpretação, demonstrando também a flexibilidade em diversas aplicações.

3.7 COMBINAÇÃO DE CLASSIFICADORES

Um procedimento de classificação ou classificador é algum método formal capaz de decidir, com base em informações fornecidas, a que grupo ou população um determinado objeto pertence. Um sistema de classificação de padrões se inicia com a coleta dos dados através de algum tipo de sensor capaz de fornecer uma representação do padrão por meio de medidas. Em uma etapa intermediária as características mais relevantes dos padrões são selecionadas através de algum método, possivelmente automático. Estas serão então utilizadas na definição da regra de classificação, que nos permitirá tomar decisões sobre os grupos aos quais os padrões pertencem (JAIN et al, 2000).

Há duas divisões principais de classificação: classificação supervisionada e classificação não supervisionada (às vezes na literatura estatísticas simplesmente referidas

como agrupamento) (WEBB, 2002). Na classificação supervisionada, temos um conjunto de amostras de dados (cada um consistindo de medidas em um conjunto de variáveis) com rótulos associados. Estes são utilizados como exemplos, para o tratamento do classificador. Na classificação não supervisionada, os dados não são rotulados e buscamos encontrar grupos na os dados e as características que distinguem um grupo a partir de uma outra (WEBB, 2002).

Formalmente, a definição de um classificador pode ser dada como uma função cujo domínio é um conjunto de variáveis (recursos) em R^n e seu co-domínio é um conjunto de rótulos de classe Ω composto de K elementos, onde $\Omega = \{ \phi_1, \dots, \phi_k \}$ (KUNCHEVA, 2004).

Na literatura são encontrados três tipos de co-domínio:

- 1) $L = \{0,1\}$, ou seja, L aborda apenas duas classes ou os rótulos significativos / não significativos para cada classificador;
- 2) $L = \{\alpha_1, \dots, \alpha_k\}$, com $k \geq 2$ fornece níveis de significância estatística para cada classificador, ou seja, um conjunto de níveis de significância com elementos $k \in Z$, onde Z é o conjunto de inteiros e w_k são rótulos simbólicos;
- 3) $L = \{S_1, S_2, \dots, S_k\}$, onde $S_i, i = 1, \dots, k$ pode ser a probabilidade, nível de confiança ou pontuação, para cada classificador.

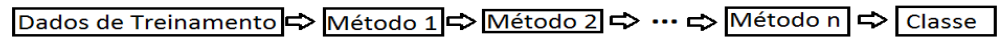
Quando se tem um problema de aplicação específico, cada classificador pode atingir graus de sucesso diferentes, sendo que nenhum deles é totalmente perfeito. Com isso, é necessária a integração de resultados de diferentes classificadores para que um melhor resultado seja obtido. Pensando nisso, o objetivo da combinação de classificadores é melhorar a eficiência na tomada de decisão ao desenvolver sistemas mais confiáveis para o reconhecimento de padrões, e melhorando o desempenho dos classificadores individuais (XU; KRZYZAK; SUEN, 1992; KUNCHEVA, 2004; DAI; LIU, 2011).

A construção de um esquema de combinação de classificadores, parte de três questões principais: escolha da arquitetura, escolha de uma estratégia de combinação e escolha do combinador adequado. Estas questões estão descritas a seguir:

- a) **Escolha da arquitetura:** Existem diversos tipos de arquiteturas na qual o classificadores são organizados, podendo ser divididas em: sequenciais, paralelas e hierárquicas (JAIN; DUIN; MAO, 2000; MORAES; MACHADO, 2012);

1. **Arquitetura sequencial:** ocorre quando cada classificador melhora as saídas dos classificadores anteriores (Figura 3).

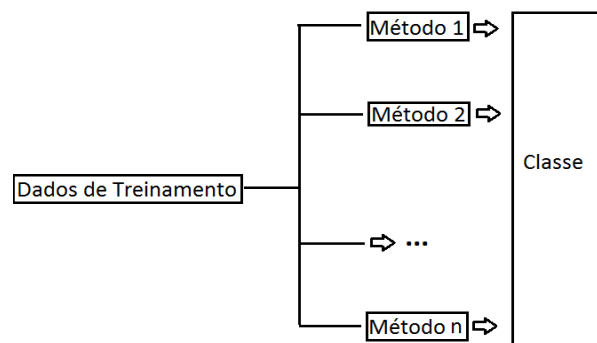
Figura 3 - Arquitetura Sequencial



Fonte: Adaptado de Moraes; Machado, 2012.

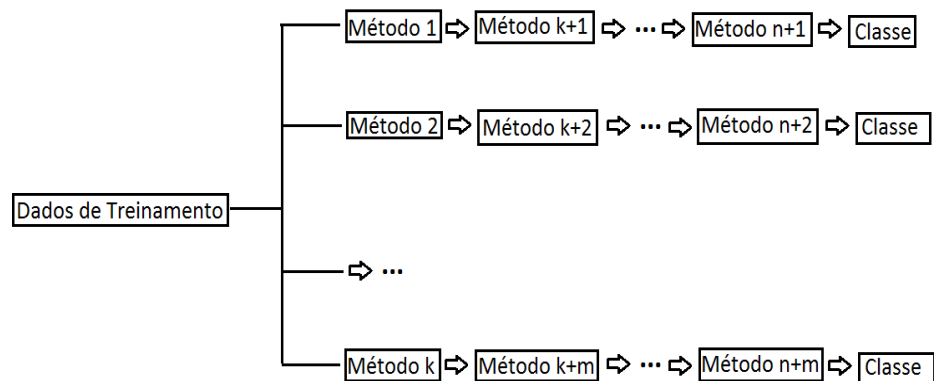
- ii. **Arquitetura paralela:** ocorre quando as saídas dos classificadores são combinadas para fornecer uma saída final (Figura 4).

Figura 4 - Arquitetura Paralela



Fonte: Adaptado de Moraes; Machado, 2012.

- iii. **Arquitetura hierárquica:** ocorre quando as saídas das arquiteturas anteriores podem ser combinadas para gerar uma saída final (Figura 5).

Figura 5 - Arquitetura Hierárquica

Fonte: Adaptado de Moraes; Machado, 2012.

- b) Escolha da estratégia de combinação:** existem dois tipos de estratégias que podem ser escolhidas, fusão que ocorre quando cada classificador escolhido deve ter acesso a exemplos para todas as variáveis envolvidas no processo ou seleção onde cada classificador escolhido precisa ter acesso a exemplos para parte das variáveis envolvido no processo e lidará apenas com este (KUNCHEVA, 2004).];
- c) Escolha do combinador adequado:** A escolha do combinador adequado baseia-se nas características que estão descritas a seguir:
- **Combinador Estático:** Executa a combinação usando uma regra pré-definida e nenhum treinamento específico é necessário (DAI;LIU, 2011; MORAES; MACHADO, 2012; DATTA;PIHUR;DATTA, 2010);
 - **Combinador Treinável:** A combinação é executada utilizando um treinamento adicional (BREVE; PONTI; MASCARENHAS, 2005; BENMOKHTAR; HUET, 2007).

Posterior a implementação e avaliação do esquema de combinação, faz-se necessária uma análise para identificar melhorias na precisão da classificação.

Considerando que a combinação de classificadores pode ser vista como uma composição das funções $g_3 \circ g_2 \circ g_1$. Um classificador é uma função g_1 dos dados da matriz X , onde $X \in R^n$ e $X = \{X_1, X_2, \dots, X_n\}$, para um conjunto Y (Figura 6). Vale salientar, que as possíveis formas de combinar as saídas dos classificadores num conjunto são

dependentes das informações que obtemos dos classificadores individuais, sendo as informações de saídas dos classificadores distintas em quatro tipos (KUNCHEVA, 2004; XU; KRZYZAK; SUEN, 1992):

- Tipo 0 (nível *Oracle*) cada classificador gera apenas duas saídas, ou seja, verdadeiro/ falso. Formalmente, temos que o vetor de saída de um classificador pode ser representado por $Y = \{0,1\}$;
- Tipo 1 (nível *Abstract*): cada classificador produz um rótulo de classe para cada X , então podemos representar como $Y = \{\phi_1, \dots, \phi_k\}$, em que $k \geq 2$. Portanto, este nível é considerado o mais universal.
- Tipo 2 (Nível *Rank*) a saída de cada classificador produz uma lista com as alternativas de classes classificadas por ordem de plausibilidade de ser o rótulo correto para cada X , podendo ser representada como $Y = \{(\phi_1, \tau_1), (\phi_2, \tau_2), \dots, (\phi_k, \tau_k)\}$, onde $\tau_i, i = 1, \dots, k$ é a classificação para o rótulo ϕ_i ;
- Tipo 3 (nível *Mensuração*): cada classificador gera um vetor de pontuação, probabilidade ou nível de confiança para cada X , esta informação pode ser representada por $Y = \{s_1, s_2, \dots, s_k\}$, onde $s_i, i = 1, \dots, k$ é a pontuação para X .

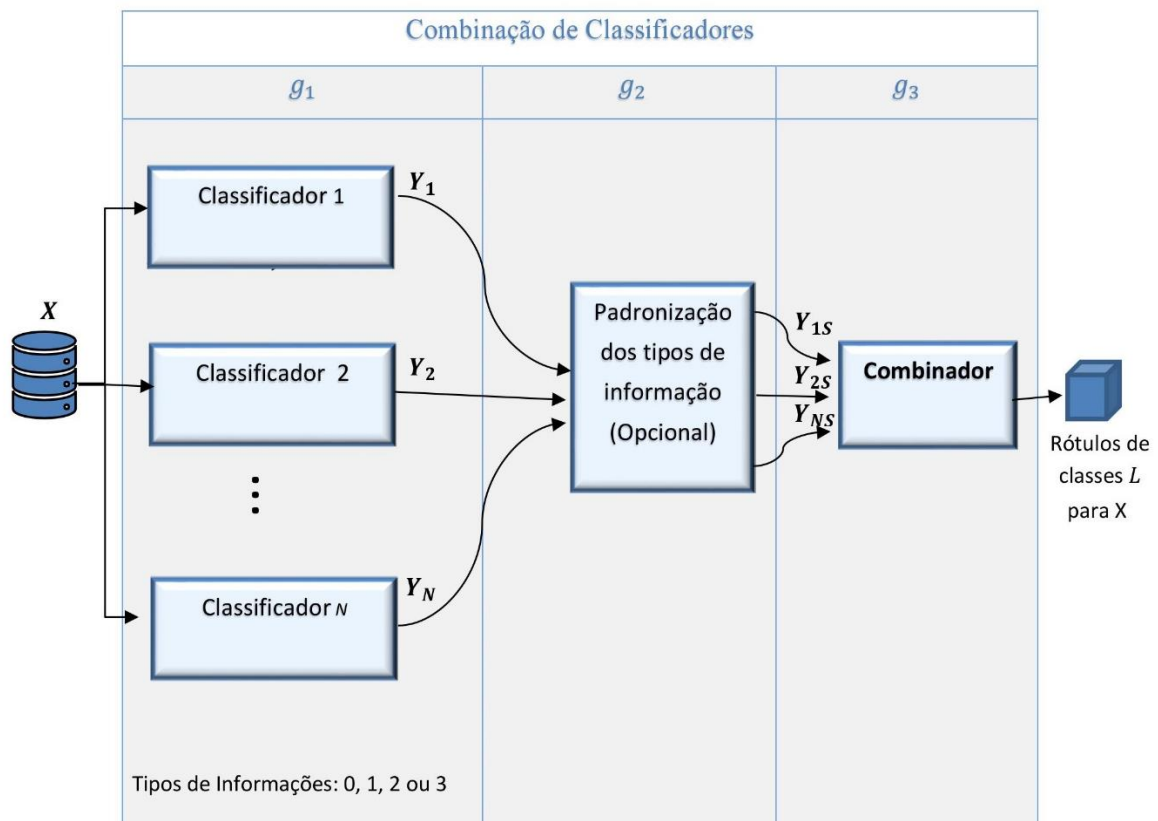
A análise do tipo de informação que cada classificador produz é fundamental para realização da combinação destes, sendo necessária a classificação de acordo com os quatro níveis descritos acima, com o intuito de detectar se todas as saídas dos classificadores no conjunto apresentam o mesmo nível de informação que a saída. Caso contrário, ou seja, se o nível de informação da saída difere entre os classificadores as vezes torna-se necessário padronizá-los previamente, para posterior aplicação do combinador.

Visando a possível necessidade de uma padronização do tipo de informação proveniente de diferentes classificadores i , uma função g_2 deve ser usada para que o nível de informação de todos os classificadores desse conjunto correspondam, para um conjunto Y_s , que é um tipo padronizado de informação, ou seja, $g_2 : Y \rightarrow Y_{is}$ para cada classificador i (Figura 6). Vale salientar, que a execução deste passo é opcional pois alguns combinadores são capazes de gerenciar diferentes tipos de informações.

O combinador para as informações saída dos classificadores pode ser um combinador fixo ou treinável, como já mostrado anteriormente na seção combinação de classificadores. A terceira função g_3 do conjunto Y_{is} (um dos tipos definidos acima) para classificação final, ou

seja, o conjunto de rótulos de classe L , onde $L = \{\alpha_1, \dots, \alpha_k\}$, onde $k \geq 2$ (Figura 6). Assim, a composição das funções $g_3 \circ g_2 \circ g_1$ corresponde para os classificadores de dados de combinação, onde a função g_2 é opcional, e depende do combinador que é usado no conjunto.

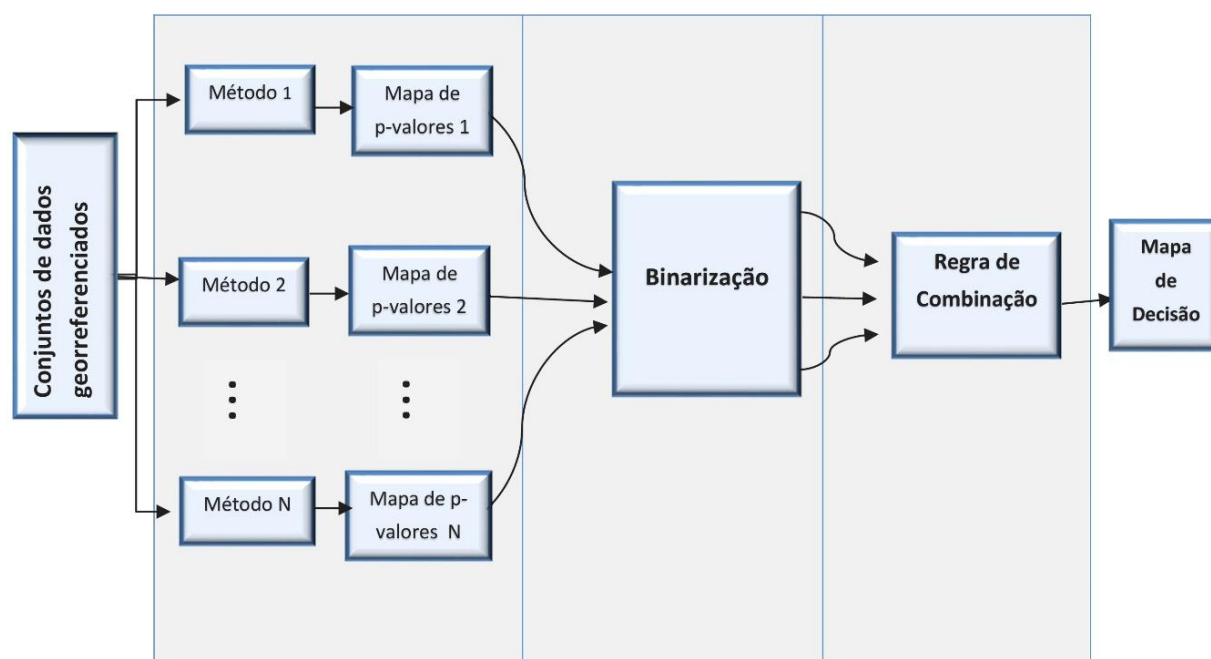
Figura 6 - Estrutura geral para combinação de classificadores



Fonte: Elaborada pelo próprio autor, 2019

Tamndo como base a combinação de classificadores, os primeiros trabalhos sobre a combinação de métodos de aglomeração espacial encontrados na literatura foram publicados por Holmes, Moraes e Vianna (2015), Holmes (2015) e Holmes, Moraes e Vianna (2018) sendo considerado um avanço na área da epidemiologia espacial, apontando para resultados melhores a partir da combinação, permitindo uma tomada de decisão mais segura. A arquitetura de combinação que foi proposta está apresentada na figura 7, onde usaram um esquema baseado em uma arquitetura paralela, onde os métodos de aglomeração espacial foram combinados por estratégia de fusão, a partir das regras de combinação *majority voting* e *plurality voting*, que são combinadores fixos.

Figura 7 - Arquitetura para a combinação de métodos de aglomeração espacial



Fonte: Figura adaptada de Holmes, 2019

Analisando mais profundamente a arquitetura, tem-se que a entrada consistiu em um conjunto de dados georreferenciados D que continham informações sobre as coordenadas dos geo-objetos z que compuseram a região geográfica, somado a informações do conjunto que seria a população R e número de casos por geo-objeto. Então: $f: G \rightarrow D$ é um mapeamento do conjunto de dados sobre as coordenadas geográficas, ou georreferenciados. Cada elemento de f pode ser denotado pelo par ordenado $c = (z, f(z))$, onde z pertence a G e $f(z)$ pertence a D . Cada um dos k métodos, onde, $k=1, \dots, N$, de aglomeração espacial (Getis Ord, Estatística Scan, Método Besag e Newell, Estatística M, Método de Tango, etc...) utiliza a estrutura f como entrada para produzir k saídas que são mapeamentos dos valores-p resultantes sobre as coordenadas geográficas $p_k: G \rightarrow [0,1]$, onde $k=1,2,\dots,N$, gerando um mapa de valores-p. Cada elemento de p_k pode ser denotado pelo par $c' = (z, p_k(c))$, onde $p_k(c)$ pertence ao intervalo $[0,1]$. Um método de binarização que transforma os valores-p em dois rótulos, onde 1 é significativo e 0 é não significativo é aplicado sobre cada elemento da estrutura p_k , onde o resultado é uma nova estrutura dada pela função $t_k: G \rightarrow \{0,1\}$. Cada elemento de t_k pode ser denotado pelo par $c'' = (z, t_k(c'))$, onde $t_k(c')$ pertence ao conjunto binário $\{0,1\}$. Finalmente, aplicou-se a regra de combinação que analisou o resultado dos k métodos anteriores, produzindo uma decisão para cada geo-objeto pertencente a estrutura de

entrada. Então foi obtida uma decisão binarizada do geo-objeto, gerando um mapa de decisão final (HOLMES, 2015).

Uma limitação encontrada na arquitetura proposta por Holmes (2015) foi a utilização da informação binária sobre valores p significativos e não significativos resultantes de métodos de agrupamento espacial, gerando uma perda de informação.

Em estudo mais recente Sá et al., (2018) utilizou uma generalização da arquitetura para identificação de municípios “prioritários”, “de transição” (que podem se tornar prioridade ou não) e “não prioritários no estado da Paraíba no tocante à dengue utilizando dados de 2011, realizando a combinação dos métodos de aglomeração espacial através da regra de votação média ponderada gerando um mapa de suporte a decisão *fuzzificado*. Este trabalho apontou um avanço ao incluir informações não binarizadas, trazendo um ganho de informações anteriormente perdidas.

3.8 REGRAS DE VOTAÇÃO E AGREGADORES COMO MÉTODO PARA COMBINAÇÃO DE CLASSIFICADORES

3.8.1 Votação por maioria

A regra que permite a combinação de diversos classificadores mais conhecida e utilizada é a votação por maioria. Esta regra de votação define como a classe vencedora aquela que obteve mais do que a metade do número total dos votos (DUIN; TAX, 2000). Então define-se que a classe que tiver mais da metade dos votos é a classe vencedora. Se não houver nenhuma classe nessa condição, então $x \in X$, logo, não recebe o rótulo, então rejeita-se (ZHOU, 2012; KUNCHEVA, 2004). Seja H uma função de decisão que registra a classe x , então:

$$H(x) = classe(j) \text{ se } \sum_{i=1}^C \Delta_{ji}(x) > \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^C \Delta_{ki}(x) \quad (11)$$

onde, $\Delta_{ji} \in 0,1$ é o voto para a classe j registrado pelo classificador i , em que K é o número de classes em Ω e C é o número de classificadores

Sendo assim, a votação é válida quando o número de votos obtidos por uma classe é superior 50% da soma de todos os votos.

3.8.2 Plurality Voting

O *plurality voting* é similar à votação por maioria quando a classificação é binária, produzindo resultados iguais (KUNCHEVA, 2004; ZHOU, 2012). Assim, $\Delta_{ji} \in 0,1$ é o voto para a classe j registrado pelo classificador i , em que K é o número de classes em Ω e C é o número de classificadores. Seja H uma função de decisão que registra a classe x , então:

$$H(x) = classe(j) se \sum_{i=1}^C \Delta_{ji}(x) = \max_{k \in K} \sum_{i=1}^C \Delta_{ki}(x) \quad (12)$$

onde, $\Delta_{ji} \in 0,1$ é o voto para a classe j registrado pelo classificador i , em que K é o número de classes em Ω e C é o número de classificadores.

Nesse caso, a classe vencedora é aquela que obtiver o maior número de votos, não sendo necessário obter mais de 50% dos votos.

3.8.3 Votação Ponderada

A votação ponderada pode ser uma opção para quando os classificadores podem ter desempenhos diferentes, pois torna possível quantificar essas diferenças, ou seja, atribuindo pesos diferentes w_i a cada classificador i de acordo com seu desempenho, podendo ser definida pela equação:

$$H(x) = classe(j) se \sum_{i=1}^C w_i \Delta_{ji}(x) = \max_{k \in K} \sum_{i=1}^C w_i \Delta_{ki}(x) \quad (13)$$

Nas aplicações práticas, os pesos são normalmente normalizados e limitados por $w_i \geq 0$ e $\sum_{i=1}^C w_i = 1$ (ZHOU, 2012).

3.8.4 Votação flexível

As votações apresentadas anteriormente podem ser utilizadas para classificadores que produzem rótulos de classes *crisp*, já para classificadores individuais que produzem saídas de probabilidade de classe, o voto flexível é geralmente a escolha. Portanto, sistemas de votação flexíveis podem ser criados ao usar probabilidades de classes para cada classificador i . Neste caso $\Delta_{ji}^S \in [0, 1]$ e pode ser obtido usando posterior probabilidades para $P(j|x)$ (ZHOU, 2012).

3.8.5 Votação Média Ponderada

O algoritmo proposto por Latif (2004) que utiliza a votação média ponderada com base em um limiar suave, traz benefícios pois usa dois parâmetros ajustáveis, cada um com uma interpretação pronta, para fornecer um desempenho de voto flexível ao usar o eleitor em aplicações diferentes. A técnica de atribuição de peso é transparente para o usuário porque o impacto do grau de concordância entre qualquer entrada de eleitor e as outras entradas é refletido diretamente no valor de peso atribuído a essa entrada. O eleitor pode ser ajustado para se comportar seja como a maioria inexata bem conhecida de votação que é geralmente usada em sistemas de controle críticos de segurança em diferentes planos de votação ou como uma votação média simples usada em muitos planos de votação sensorial, dando melhor desempenho em termos de segurança e disponibilidade.

O eleitor médio ponderado fornece a média ponderada dos valores obtidos dos módulos redundantes. Dado um conjunto de entradas (resultados do módulo redundante) $x_1; x_2$ e x_3 para um determinado ciclo de votação, um eleitor médio ponderado baseado na medição de distância determina a distância numérica dos pares de entrada, $d_{12} = |x_1 - x_2|$; $d_{13} = |x_1 - x_3|$ e $d_{32} = |x_3 - x_2|$ partir da qual são calculados valores de ponderação de entradas individuais, $w_1; w_2$ e w_3 . Os valores de peso são então usados para calcular um único valor como a saída do eleitor (LATIF, 2004). A determinação do grau de proximidade de todas as entradas em pares da votação. Para todos os pares de entrada votação i e j ; o grau de s_{ij} proximidade é definido como se segue:

$$s_{ij} = \begin{cases} 1 & \text{se } d_{ij} \leq a \\ \left(\frac{n}{n-1}\right) \left(1 - \frac{d_{ij}}{n \cdot a}\right) & \text{se } a < d_{ij} < n \cdot a \\ 0 & \text{se } d_{ij} \geq n \cdot a \end{cases} \quad (15)$$

onde d_{ij} é a distância numérica dos pares de entrada i e j ; a é o valor de limiar fixo de votação (utilizado em eleitores convencionais de maioria inexatas), e n é um sintonizável constante de proporcionalidade parâmetro a ser utilizado para controlar a votação. Chamamos s_{ij} como o indicador acordo com os pares de entrada x_i e x_j : Para pares de entrada com uma distância menor do que uma aplicação específica- um limiar; $s_{ij}=1$; e para aqueles com uma distância maior que $n \cdot a$ (onde n é uma constante de proporcionalidade) $s_{ij}=0$: pares de entrada com uma distância entre a e $n \cdot a$ um é dado um valor a partir do intervalo $[0 \ 1]$ para s_{ij} .

De acordo, com os valores dos indicadores computados para todos os pares de entrada eleitor, o valor ponderado de cada i eleitor é definida com base na seguinte equação (m - entrada assumidas).

$$w_i = \frac{\sum_{j=1, j \neq i}^m s_{ij}}{m-1} \mid i, j = 1, \dots, m \text{ e sendo } i = j: s_{ij}=0 \quad (16)$$

A saída de eleitor é, então, dado pela Eq. 13:

$$y = \frac{\sum_{i=1}^m x_i \cdot w_i}{\sum_{i=1}^m w_i} \quad (17)$$

3.8.6 Cardinalidade de Conjuntos Fuzzy

A medida de cardinalidade fuzzy $\Sigma Count$ sugerida por Zadeh, generaliza a medida de contagem clássica. Essa generalização ao ser aplicada permite que muitas estruturas combinatórias e técnicas de contagem sejam fuzzificadas e portanto, usadas em modelos de representação de conhecimento e reconhecimento de padrões (KOSKO, 1986).

De acordo com Zadeh (1983), a cardinalidade de um conjunto *fuzzy* X , denotado por $\Sigma Count$, é um número real definido como a soma dos seus n valores de pertinência μ :

$$\Sigma Count(X) = \sum_{i=1}^n \mu(x_i), i = 1, \dots, n \quad (18)$$

3.8.7 $\Sigma Count$ Ponderado

A contagem $\Sigma Count$ pode ser ponderada de acordo com um vetor de pesos $w = (w_1, \dots, w_n)$, onde cada elemento é um número real não-negativo que define um peso para cada um dos n valores de pertinência μ do conjunto *fuzzy* X .

$$\Sigma Count(X; w) = \sum_{i=1}^n w_i \mu(x_i), i = 1, \dots, n, \text{ onde } \sum_{i=1}^n w_i = 1. \quad (19)$$

Apesar de Zadeh (1983) não ter colocado explicitamente uma restrição sobre a soma dos pesos, por uma questão de compatibilidade com sua contraparte clássica, imporemos essa restrição.

3.8.8 Funções H_θ

As funções H_θ , são uma relevante classe de funções de mistura generalizadas do tipo de média ponderada em que os pesos não são fixos, que foram aplicadas em estudo realizado por Farias (2018) como um mecanismo para redução de imagens e um método para combinar múltiplos algoritmos de comitês classificadores.

Nesse sentido as funções H_θ , foram utilizadas como combinador de classificadores, estas são construídas a partir de função $\theta: [0,1]^n \rightarrow [0,1]$ (FARIAS; SANTIAGO; BEDREGAL, 2016), conforme definida na equação a seguir:

Seja $\theta: [0,1]^n \rightarrow [0,1]$ uma função.

Então,

$$H_{\theta}(x) = \begin{cases} x_1, & \text{se } x_1 = \dots = x_n \\ \frac{1}{n-1} \sum_{i=1}^n \left(x_i - \frac{x_i |x_i - \theta(x)|}{\sum_{j=1}^n |x_j - \theta(x)|} \right) & , \text{ caso contrário} \end{cases} \quad (20)$$

onde $X = (x_1, \dots, x_n) \in [0, 1]^n$ é um vetor de entrada n-dimensional, sendo um vetor da função de mistura generalizada.

Assim, Farias, et al. (2016) estabeleceu uma família de funções de misturas generalizadas utilizando as funções de média ponderada ordenada (OWA), que foi introduzido por Ronald R. Yager (1988), ou seja, formaram uma nova classe de regras de combinação baseada na função $H_{\theta}(x)$ usando funções Mínimo, Máximo, Aritmético, Média, mediana e cOWA. tais como: H_{Med} , H_{Arith} , H_{Max} , H_{Min} e H_{cOWA} . Então, nos casos em que θ é *Max* ou *Med* ou *Min* ou *cOWA* ou *Arith*, temos:

$$H_{Med}(x) = \begin{cases} x, & \text{se } X = (x, \dots, x) \\ \frac{1}{n-1} \sum_{i=1}^n \left(x_i - \frac{x_i |x_i - Med(x)|}{\sum_{j=1}^n |x_j - Med(x)|} \right) & , \text{ caso contrário} \end{cases} \quad (21)$$

$$H_{Min}(x) = \begin{cases} x_1, & \text{se } x_1 = \dots = x_n \\ \frac{1}{n-1} \sum_{i=1}^n \left(x_i - \frac{x_i |x_i - Min(x)|}{\sum_{j=1}^n |x_j - Min(x)|} \right) & , \text{ caso contrário} \end{cases} \quad (22)$$

$$H_{Max}(x) = \begin{cases} x_1, & \text{se } x_1 = \dots = x_n \\ \frac{1}{n-1} \sum_{i=1}^n \left(x_i - \frac{x_i |x_i - Max(x)|}{\sum_{j=1}^n |x_j - Max(x)|} \right) & , \text{ caso contrário} \end{cases} \quad (23)$$

$$H_{Arith}(x) = \begin{cases} x_1, & \text{se } x_1 = \dots = x_n \\ \frac{1}{n-1} \sum_{i=1}^n \left(x_i - \frac{x_i |x_i - Arith(x)|}{\sum_{j=1}^n |x_j - Arith(x)|} \right) & , \text{ caso contrário} \end{cases} \quad (24)$$

$$H_{cOWA}(x) = \begin{cases} x_1, & \text{se } x_1 = \dots = x_n \\ \frac{1}{n-1} \sum_{i=1}^n \left(x_i - \frac{x_i |x_i - cOWA(x)|}{\sum_{j=1}^n |x_j - cOWA(x)|} \right) & , \text{caso contrário} \end{cases} \quad (25)$$

Essas funções foram utilizadas de forma inédita como regra de combinação para métodos de aglomeração espacial no segundo estudo de caso desta tese, onde serão apresentados os resultados da aplicação de cada função (subseção 6.2).

“Ensinar não é transferir conhecimento, mas criar as possibilidades para a sua produção ou a sua construção.”

Paulo Freire

4 NOVA ARQUITETURA UTILIZANDO REGRAS DE COMBINAÇÃO FUZZY PARA MÉTODOS DE AGLOMERAÇÃO ESPACIAL APLICADA À EPIDEMIOLOGIA

A arquitetura encontrada na literatura para a combinação de métodos de aglomeração espacial proposta por Holmes (2015) permitiu uma flexibilidade para o uso de qualquer método de aglomeração espacial e diferentes regras de combinação, porém se faz necessário a binarização dos dados, como já discutido anteriormente.

A nova proposta de arquitetura para combinar métodos de aglomeração espacial utilizando regras de combinação *fuzzy* permite ter como entrada dados georreferenciados, adotando um processo de fuzzificação para inclusão das incertezas, com os diferentes níveis de significância de cada localidade, sendo construídas as funções de pertinência, a fim de obter informações mais ampliadas do fenômeno estudado, além de utilizar um combinador ajustável, diferenciando-se do proposto anteriormente.

Partindo do entendimento da estrutura geral dos métodos de aglomeração espacial (Figura 1) e da combinação de classificadores (figura 6) mostrada anteriormente, iniciaremos a construção da arquitetura para combinação de N métodos de aglomeração espacial.

Retomando, um método de aglomeração espacial pode ser visto como uma composição de funções diferentes. Nesse caso, a informação é georreferenciada; então a função f_1 é definida a partir do domínio de geo-objetos na região geográfica $G \in R_+^2$ e um conjunto de dados georreferenciados D relacionado a cada geo-objeto. Então podemos considerar, $D \in R^n$ e $D = \{D_1, D_2, \dots, D_n\}$, então o domínio da função f_1 é D^G e um método aglomeração espacial como definido anteriormente, é um mapeamento $f_1 : D^G \rightarrow \Omega^G$, onde $\Omega \in [0,1]$, é o p -valor calculado pelo método de aglomeração espacial para o geo-objeto. A segunda função é um mapeamento $f_2 : \Omega^G \rightarrow L^G$, que fornece um mapa cujo conjunto de rótulos de classe é L^G , em que $L^G = \{\phi_1, \dots, \phi_k\}$ para cada objeto geográfico em G , onde L^G tem as características mencionadas acima, conforme apresentado na figura 7. Portanto, é uma composição de funções, que pode ser denotada como $f_2 \circ f_1$.

A fim de combinar os métodos de aglomeração espacial (Figura 8), inicialmente analisa-se o tipo de informação de saída de acordo com os quatro níveis especificados anteriormente. De acordo com as características da função f_2 , ela pode fornecer três níveis de informação para co-domínio L :

- Tipo 0 (nível *Oracle*): cada método de aglomeração espacial gera apenas duas classes "significativo" e "não significativo" para cada geo-objeto, ou seja, $L^G = \{0,1\}$;
- Tipo 1 (nível *Abstract*): cada método de aglomeração espacial gera um rótulo de classe para cada geo-objeto, ou seja, $L^G = \{\phi_1, \dots, \phi_k\}$, com $k \geq 2$;
- Tipo 3 (nível *Mensuração*): cada método de aglomeração espacial produz uma pontuação, probabilidade ou nível de confiança para cada objeto geográfico, ou seja, $L^G = \{S_1, S_2, \dots, S_k\}$, onde $S_i, i = 1, \dots, k$.

O nível do tipo 2 ou *Rank* não foi identificado na literatura científica para métodos de aglomeração espacial. Os níveis 0, 1 e 3 são capazes de fornecerem um mapa geográfico com diferentes rótulos de classe, simbólicos ou numéricos.

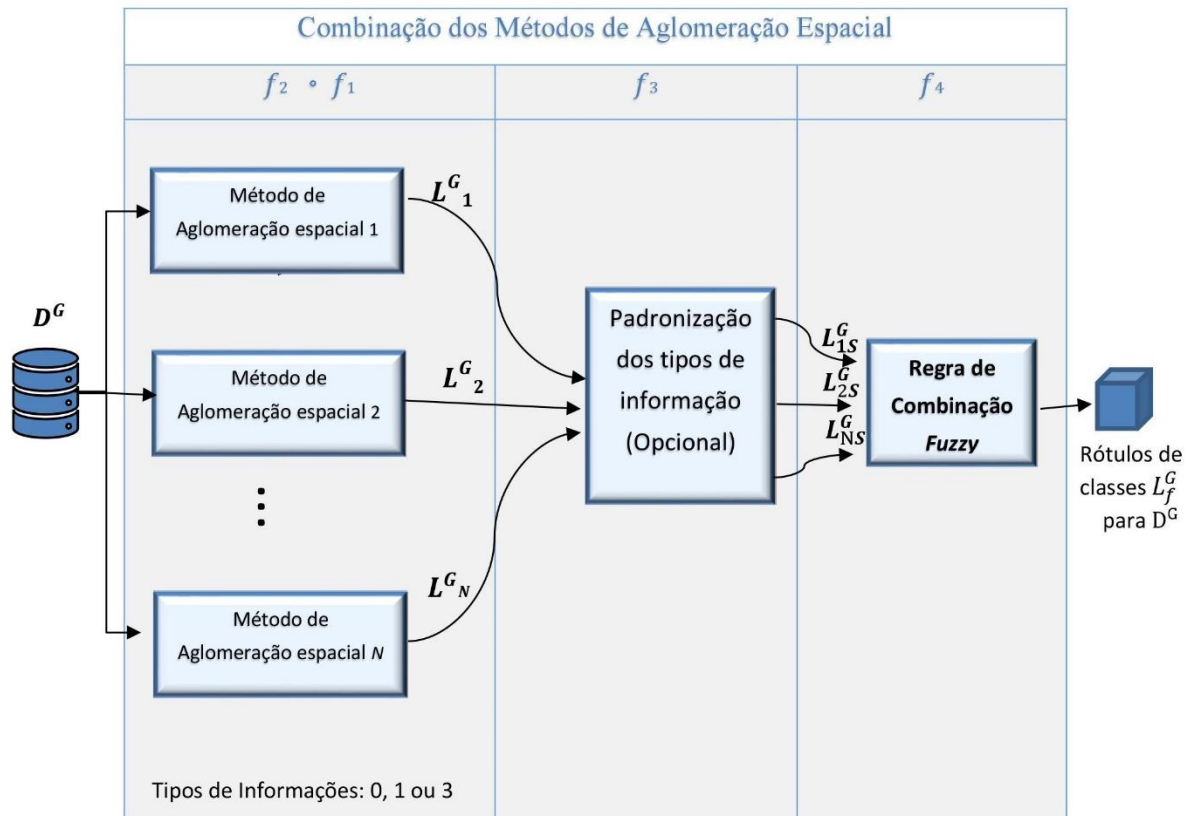
A depender do tipo de combinador escolhido, pode ser preciso padronizar os níveis de informações de saída dos métodos de aglomeração espacial, quando estes não tiverem os mesmos níveis. Quando for necessário padronizar o tipo de informação que vem de diferentes métodos de aglomeração espacial (Figura 8), uma função f_3 deve ser usada para que os tipos de informações estejam padronizadas para todos os métodos, usando o conjunto com nível mais baixo encontrado, que é o conjunto L_s . Uma função binária pode funcionar para todos os casos, padronizando o nível de informação para o Tipo 0 de todos os métodos de aglomeração presentes no conjunto, isto é, neste caso, $f_3 : L^G \rightarrow L_s^G$, onde $L_s^G = \{0,1\}$. Contudo, esta binarização pode não ser interessante devido a perda de informações que ocorre. Então, podemos adotar uma função de *fuzzificação* em f_3 , gerando informação de saída do Tipo 3 para todos os métodos de aglomeração espacial presentes no conjunto, neste caso, $f_3 : L^G \rightarrow L_s^G$, onde $L_s^G = \{(1 - S_1), (1 - S_2), \dots, (1 - S_k)\}$, ou seja, os graus de pertinência intervalados em $[0,1]$. É importante ressaltar que este passo é opcional, como na combinação de classificadores e pelas mesmas razões.

A quarta função f_4 consiste na aplicação do combinador, este pode ser fixo ou treinável. Então, a função f_4 do conjunto L_s para o mapa de decisão final, ou seja, o conjunto de rótulos L_f , onde L_f pode fornecer três tipos de informação, de acordo com o domínio da função f_4 (Figura 8). Por exemplo, se domínio f_4 é $\{0,1\}$, seu co-domínio pode ser rotulado como $L_f = \{0,1\}$, ou simbólicos, como $L_f^G = \{\text{significativa}, \text{não significativa}\}$, ou $L_f^G = \{\text{prioritário}, \text{não prioritário}\}$. Como também, se domínio f_4 é $[0,1]$, seu co-domínio pode ser rotulado como $L_f^G = [0,1]$, ou simbólico, como $L_f^G = \{\text{prioritário}, \text{em transição}, \text{não prioritário}\}$.

Portanto, a composição das funções $f_4 \circ f_3 \circ f_2 \circ f_1$ corresponde à combinação

dos métodos de aglomeração espacial. Como explicado acima, a função f_3 pode ser necessária ou não de acordo com os tipos de informação de todos os métodos presentes no conjunto e o combinador.

Figura 8 – Nova arquitetura utilizando regras de combinação *fuzzy* para métodos de aglomeração espacial



Fonte: Elaborada pelo próprio autor, 2019.

Ao compararmos esta nova arquitetura com a proposta anteriormente por Holmes (Figura 7), observamos que a parte inicial que corresponde as funções $f_2 \circ f_1$, são idênticas nas duas arquiteturas. A partir das funções f_3 e f_4 , as arquiteturas são diferentes, pois na nova proposta as funções são baseadas em *fuzzy*, não necessitando da binarização que ocorria obrigatoriamente na anterior. Então, esta nova proposta parte da ideia que a combinação dos métodos de aglomeração espacial com a inclusão das incertezas presentes no processo saúde-doença, produz melhores resultados do que se utilizados individualmente, tornando possível a obtenção de informações mais fidedignas e tomada de decisões mais precisas.

Em resumo, esta nova arquitetura é do tipo paralela pois as saídas dos classificadores são combinadas para fornecer uma saída final, utilizando como estratégia fusão, permite também o uso de três níveis de informações (níveis 0, 1 e 3) e permite uma flexibilidade em

relação a utilização de qualquer combinador, sendo capaz de produzir resultados diversificados, fornecendo mapas para uma tomada de decisão embasada cientificamente.

4.1 NOVA PROPOSTA PARA REGRAS DE VOTAÇÃO UTILIZANDO CARDINALIDADE DE CONJUNTOS *FUZZY*

O aprofundamento teórico sobre regras de combinação para aplicação na nova arquitetura levou a construção de duas novas propostas de regras de votações baseadas em cardinalidade de conjuntos *Fuzzy*, que foram aplicadas no terceiro estudo de caso desta tese e estão descritas a seguir.

4.1.1 Votação por maioria *fuzzy* utilizando cardinalidade de conjuntos *fuzzy*

A regra de votação por maioria demonstrada anteriormente na equação (11) define que a classe que tiver mais da metade dos votos é a classe vencedora. A generalização desta culminou na construção da regra de votação por maioria *fuzzy* baseada na cardinalidade de conjuntos *fuzzy*, determina como classe vencedora o que obtiver o valor da função de pertinência maior que o ponto de corte determinado. A definição da regra é dada por:

$$H(x) = classe(j) \text{ se } \sum_{i=1}^C \Sigma Count_{ij}(X; w_i) > \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^C \Sigma Count_{ki}(X; w_i) \quad (26)$$

onde, K = número de classes em Ω ; C = número de classificadores; e o $\Sigma Count_{ij}(X; w_i)$ é a contagem de votos para a classe j usando os classificadores i para o vetor de dados X e o vetor de pesos w .

Portanto, fará parte da classe dos significativos o que obtiver o valor da função de pertinência maior que o ponto de corte determinado.

4.1.2 *Plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy*

O *plurality voting* já discutido anteriormente na equação (12) determina que a classe vencedora é aquela que recebe o maior número de votos, não sendo necessário conseguir obter mais de 50% dos votos dos classificadores. Segundo Zhou (2012), no caso da

classificação binária, a votação por maioria e o *plurality voting* produzem o mesmo resultado. A classe vencedora é determinada a partir do somatório das funções de pertinência das classes contra e o somatório das funções de pertinência das classes a favor, sendo vencedora a classe que apresentar o maior valor.

$$H(x) = classe(j)se \sum_{i=1}^C \Sigma Count_{ji(x)} = \max_{k \in K} \sum_{i=1}^C \Sigma Count_{ki(x)} \quad (27)$$

onde, K = número de classes em Ω ; C = número de classificadores; e o $\Sigma Count_{ij}(X; w_i)$ é a contagem de votos para a classe j usando os classificadores i para o vetor de dados X e o vetor de pesos w .

A generalização desta regra de votação, ou seja, o *plurality voting fuzzy* baseada na cardinalidade de conjuntos *fuzzy*, permitiu a utilização de informações não binarizadas trazendo resultados diferentes da votação por maioria *fuzzy* baseada na cardinalidade de conjuntos *fuzzy*.

“Ninguém ignora tudo. Ninguém sabe tudo. Todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa. Por isso aprendemos sempre.”

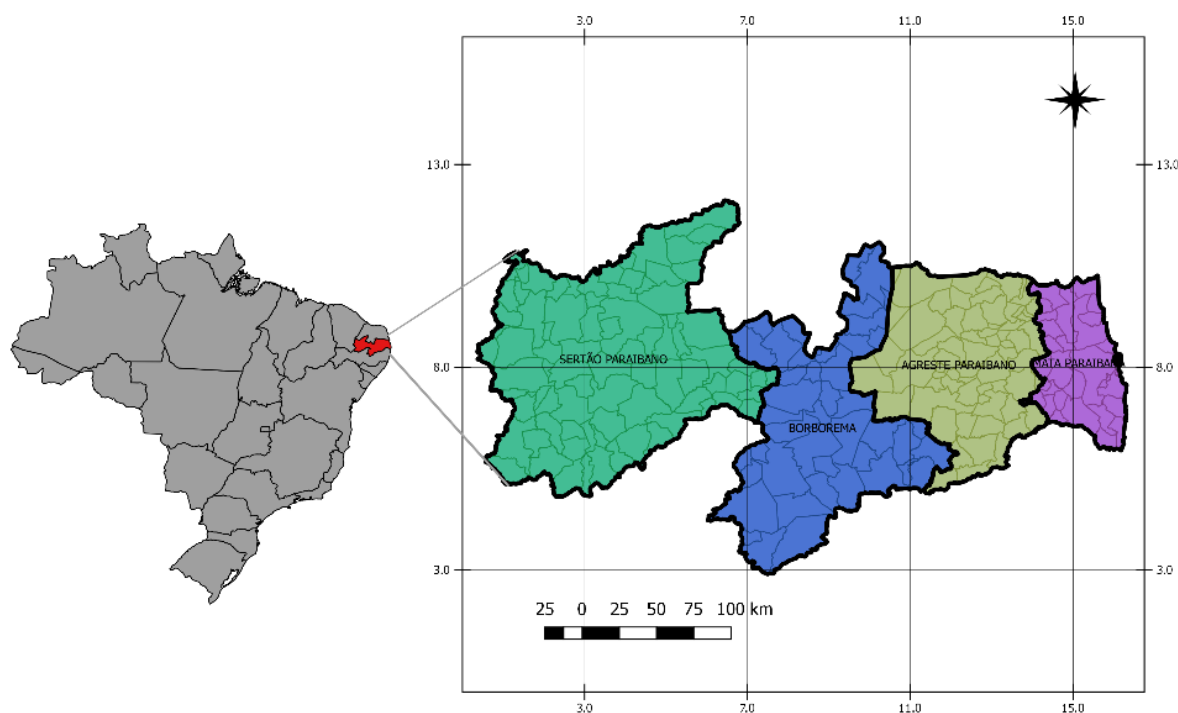
Paulo Freire

5 PERCURSO METODOLÓGICO

5.1 ÁREA DE ESTUDO

Elegeram-se como área de investigação, o estado da Paraíba, que ocupa uma extensão territorial de 56.585 km², com uma população em torno 3.943.885 habitantes, constituindo-se por 223 municípios (IBGE, 2010). Na figura 9, observa-se o mapa geográfico do estado da Paraíba.

Figura 9 - Mapa geográfico do Estado da Paraíba



Fonte: SANTOS, 2018.

5.2 CARACTERIZAÇÃO DO ESTUDO E ANÁLISE DOS DADOS

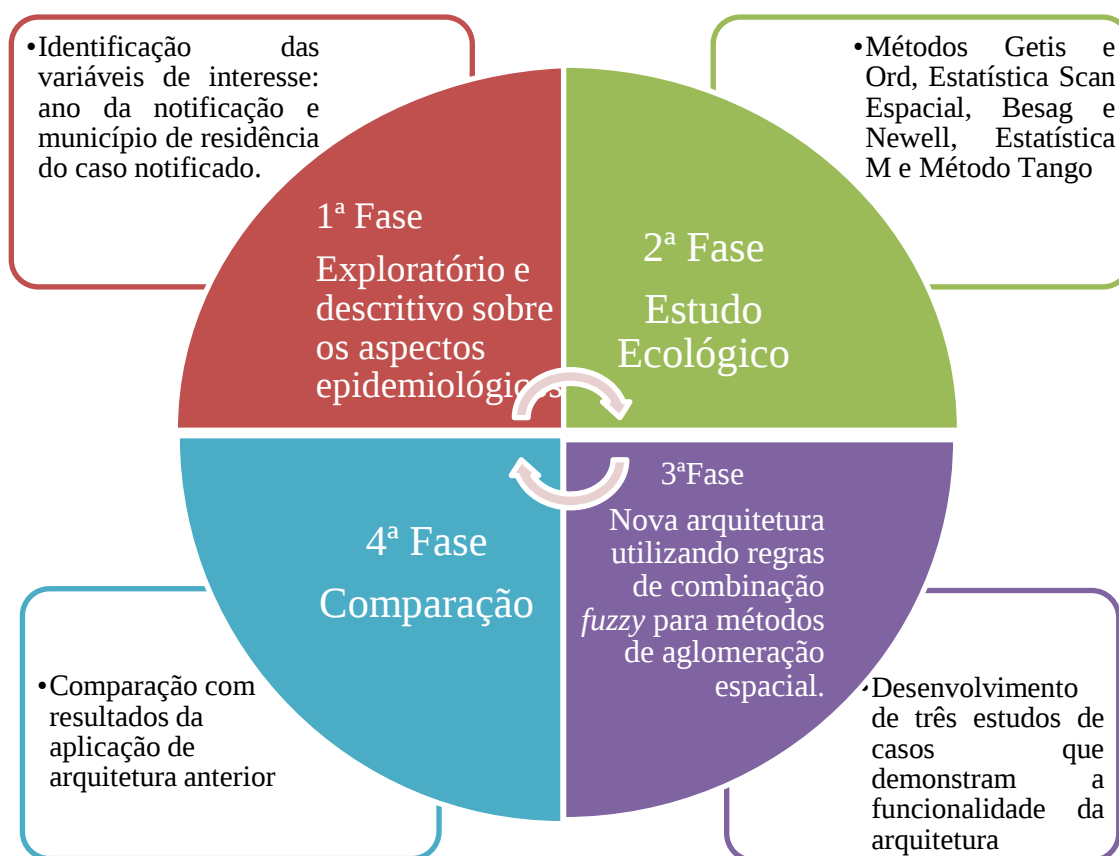
Caracteriza-se por ser um estudo ecológico de abordagem quantitativa, o qual utilizou dados do tipo secundário dos casos registrados no Sinan/Dengue da Secretaria Estadual de Saúde da Paraíba, do ano de 2011 e 2017. Em relação aos dados da população residente no

município foram obtidos a partir do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010; PNAD, 2014).

Sequencialmente foi realizado um refinamento do banco de dados para identificação das variáveis de interesse: ano da notificação e município de residência do caso notificado. Testou-se a normalidade dos dados com a aplicação do teste de *Lilliefors*, para verificar se a distribuição das variáveis pode ser aproximada pela distribuição Normal (SIEGEL, 2006).

Na área da estatística espacial foi utilizado os seguintes métodos de aglomerados espaciais: Getis e Ord, Estatística Scan Espacial, Besag e Newell, Estatística M e Método Tango. A seguir, serão demonstrados o percurso metodológico do três estudos de casos distintos usando os dados epidemiológicos dos casos de Dengue, no estado da Paraíba, Brasil, que demonstram a funcionalidade da arquitetura proposta através da combinação dos métodos utilizando diferentes regras de combinação, além da comparação com arquitetura anterior com intuito de apontar melhorias nos resultados finais.

Figura 10 – Percurso Metodológico do estudo.



Fonte: Elaborada pelo próprio autor, 2019.

5.3 ASPECTOS ÉTICOS

O projeto foi avaliado diretamente pela Secretaria de Saúde do Estado da Paraíba, Brasil, havendo liberação dos dados do tipo secundário dos casos registrados no Sinan/Dengue, respeitando assim os critérios estabelecidos na Resolução nº 466/12 de 12 de dezembro de 2012, que regulamenta a pesquisa em seres humanos.

“As raízes da educação são amargas, mas o fruto é doce.”.

Aristóteles

6 RESULTADOS

Este capítulo, contém os resultados da aplicação da arquitetura proposta para a combinação dos métodos aglomeração espacial dos três estudos de casos distintos utilizando os dados epidemiológicos dos casos de Dengue, no estado da Paraíba, Brasil.

O primeiro estudo de caso utilizou a regra de votação média ponderada e obteve um mapa final com a classificação de cada município segundo “municípios prioritários”, “municípios de transição” (que podem se tornar prioridade ou não) e “não prioritário”. O segundo estudo de caso identificou os municípios significativos e não significativos para o dengue, utilizando a classe de funções de H_0 . O último estudo de caso trouxe dados do dengue mais recentes do ano 2017, esse estudo conseguiu avançar em relação aos anteriores pois utilizou votação por maioria *fuzzy* utilizando cardinalidade de conjuntos *fuzzy* e o *plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy*, além da construção do mapa por classes.

6.1 ESTUDO DE CASO I

O primeiro estudo de caso consistiu na aplicação da arquitetura proposta para combinação de métodos de aglomeração espacial, utilizando os dados do Dengue para o ano 2011, no estado da Paraíba, Brasil (Figura 11). Na análise espacial foram escolhidos os métodos de aglomeração espacial: Estatística Scan espacial, Método Besag Newell e Getis Ord.

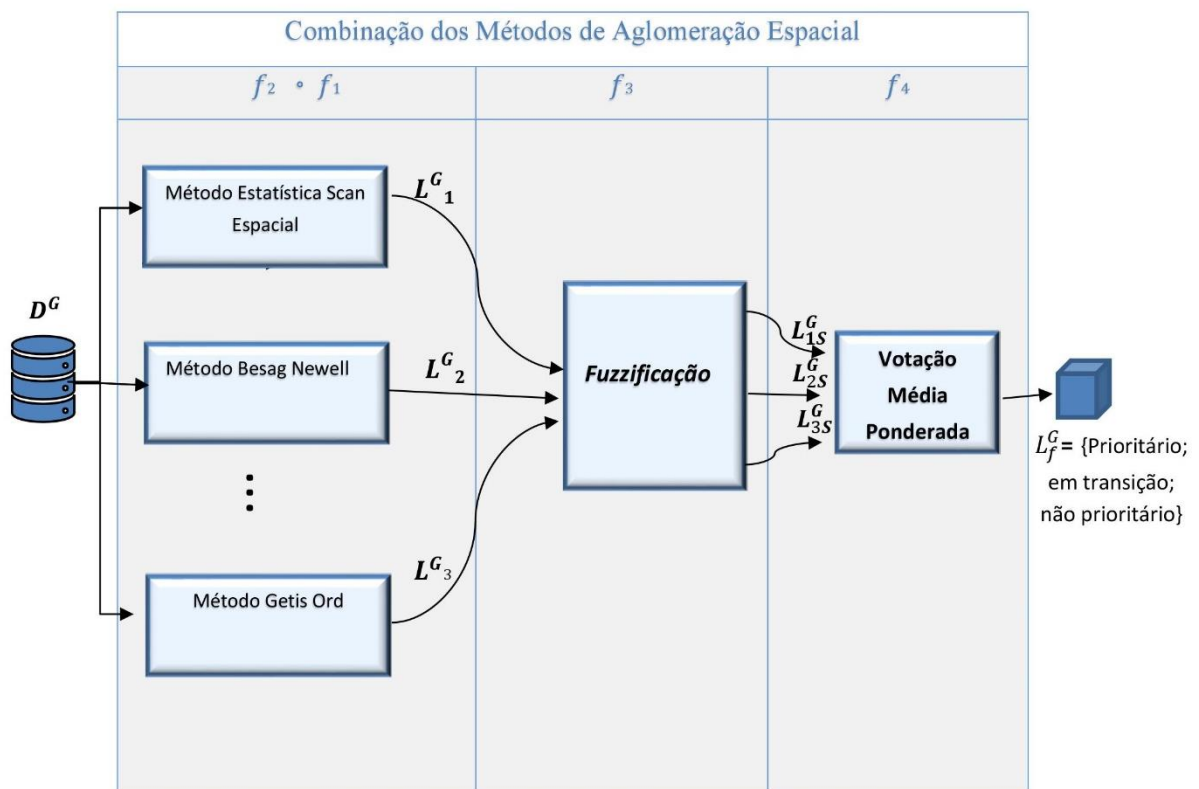
A aplicação da arquitetura inicia com a composição de funções, que pode ser denotada como $f_2 \circ f_1$. Os métodos escolhidos usaram como entrada um banco de dados georreferenciados, ou seja, a função $f_1 : D^G \rightarrow \Omega^G$, onde G é a região geográfica que contém geo-objetos z que são os municípios, D são os números de casos do Dengue e a população de cada município (geo-objeto) e onde $\Omega \in [0,1]$, é o p -valor calculado pelo método de aglomeração espacial para cada geo-objeto. A segunda função é um mapeamento $f_2 : \Omega^G \rightarrow L^G$, que fornece um mapa cujo conjunto de rótulos de classe é L^G , em que $L^G = \{s_1, \dots, s_k\}$ para cada objeto geográfico em G , nesse caso o nível de informação fornecido é tipo 3.

A função f_3 é realizada quando há necessidade de padronização da informação, embora opcional nesse caso, foi operacionalizada com intuito de realizar *fuzzificação* da informação. Então nesse caso, podemos adotamos uma função de *fuzzificação* em f_3 , gerando

informação de saída do Tipo 3 para todos os métodos de aglomeração espacial presentes no conjunto, neste caso, $f_3 : L^G \rightarrow L_S^G$, onde $L_S^G = \{(1 - S_1), (1 - S_2), \dots, (1 - S_k)\}$, ou seja, os graus de pertinência intervalados em $[0,1]$.

A quarta função f_4 consiste na aplicação do combinador, neste caso foi escolhida a regra de votação média ponderada proposta por Latif (2004), partindo da combinação dos graus de pertinência. Portanto, o domínio f_4 é $[0,1]$ e seu co-domínio pode ser definido como $L_f = \{\text{prioritário, em transição, não prioritário}\}$. Por fim, foi gerado o mapa da combinação dos classificadores onde os geo-objetos foram rotulados como prioritários, transição (municípios que podem se tornar prioritários ou estão deixando de sê-lo) e não prioritários.

Figura 11 – Aplicação da arquitetura para Combinação dos Métodos de Aglomeração Espacial utilizando o combinador Votação média ponderada para os dados do Dengue, Paraíba, Brasil

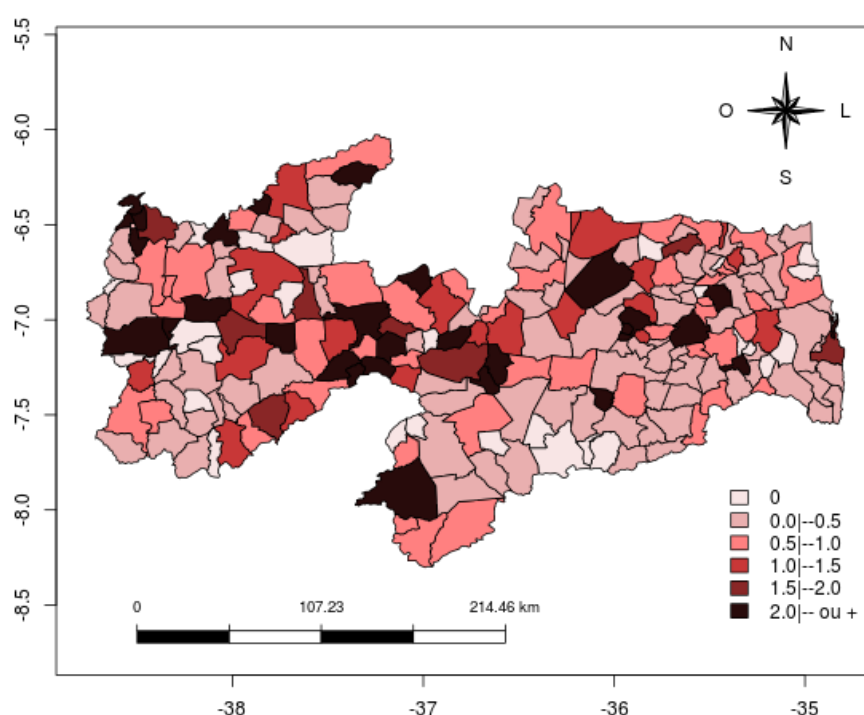


Fonte: Elaborada pelo próprio autor, 2019.

No ano de 2011, foram notificados nos 11.490 casos de Dengue em todo estado da Paraíba-Brasil (SINAN/Dengue, 2014). Devido variabilidades metodológicas de implementação dos métodos de aglomeração espacial disponíveis os resultados gerados são distintos uns dos outros. Com o intuito de verificar qual desses resultados é mais próximo da

realidade epidemiológica, o mapa do RIE (equação 10) foi gerado para representar a intensidade da ocorrência Dengue nos municípios da Paraíba para o ano de 2011, sendo utilizado como referência para comparação com os mapas dos métodos de aglomeração espacial (Figura 12). Onde o RIE apresentou variação entre 0 e 6,07 por 100.000 habitantes e mostrou uma heterogeneidade a distribuição da doença em todo estado.

Figura 12 - Mapa da Razão de Incidência espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011

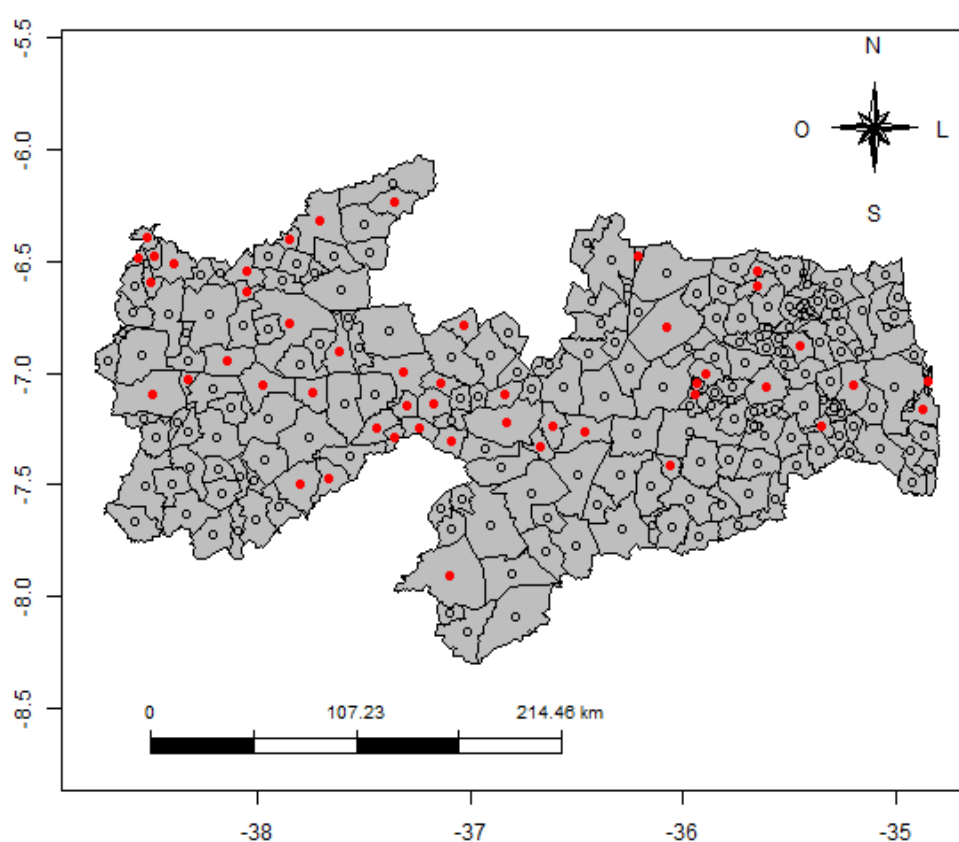


Fonte : Elaborado pelo próprio autor, 2017.

Neste estudo de caso, optou-se pelo modelo de probabilidade de Poisson para o método *Scan* espacial permitindo identificar os conglomerados espaciais. Foram testados percentuais populacionais de 0,1%, 0,5%, 0,7% e 1%, para saber qual apresentaria resultados mais satisfatórios sob o ponto de vista epidemiológico através da comparação com o mapa de RIE. O percentual mais adequado foi com restrição de 0,1% da população.

O mapa da estatística *Scan* (Figura 13) identificou conglomerados de alto e baixo risco. Quarenta e oito (48) municípios pertencem a conglomerados espaciais para o ano 2011, com percentual da população de 0,1%. Comparando-se o mapa *Scan* com o mapa de risco relativo, pode-se verificar a coincidência das regiões que apresentam os maiores índices de RIE com as que possuem conglomerados espaciais significativos.

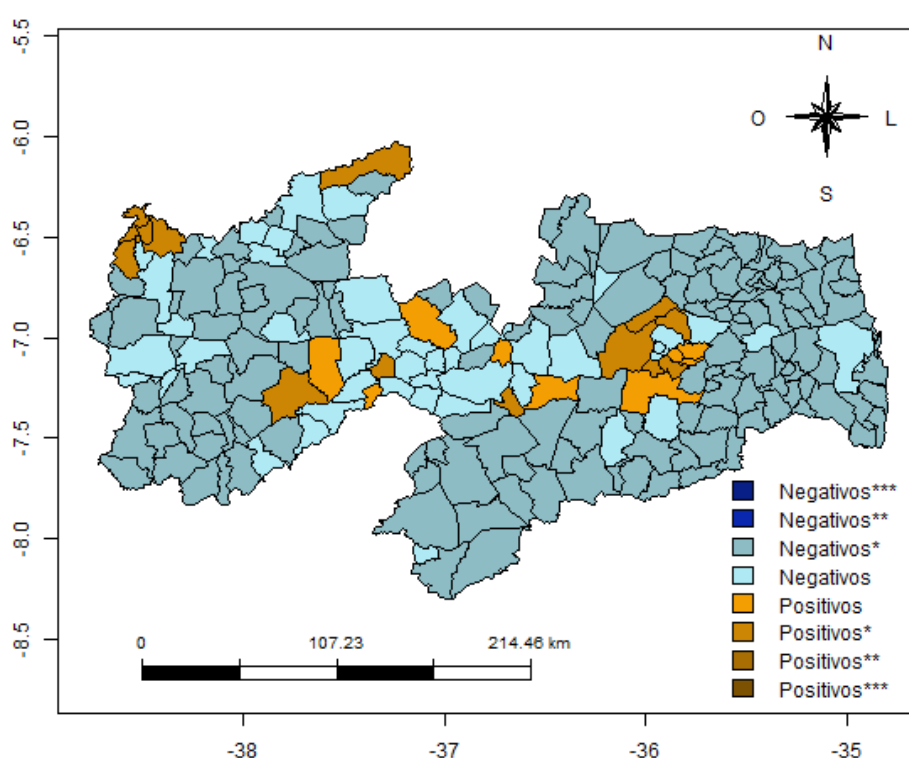
Figura 13 - Mapa dos valores-*p* do método *Scan* espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte : Elaborado pelo próprio autor, 2017.

A figura 14, mostra o mapa resultante da aplicação do método Getis Ord, observou-se que apenas alguns municípios apresentaram aglomerados espaciais de valores altos. Estes, localizaram-se em sua maioria na mesorregião do Agreste e sertão Paraibano. Os aglomerados espaciais de valores baixos estavam distribuídos de forma heterogênea em toda a região da Paraíba.

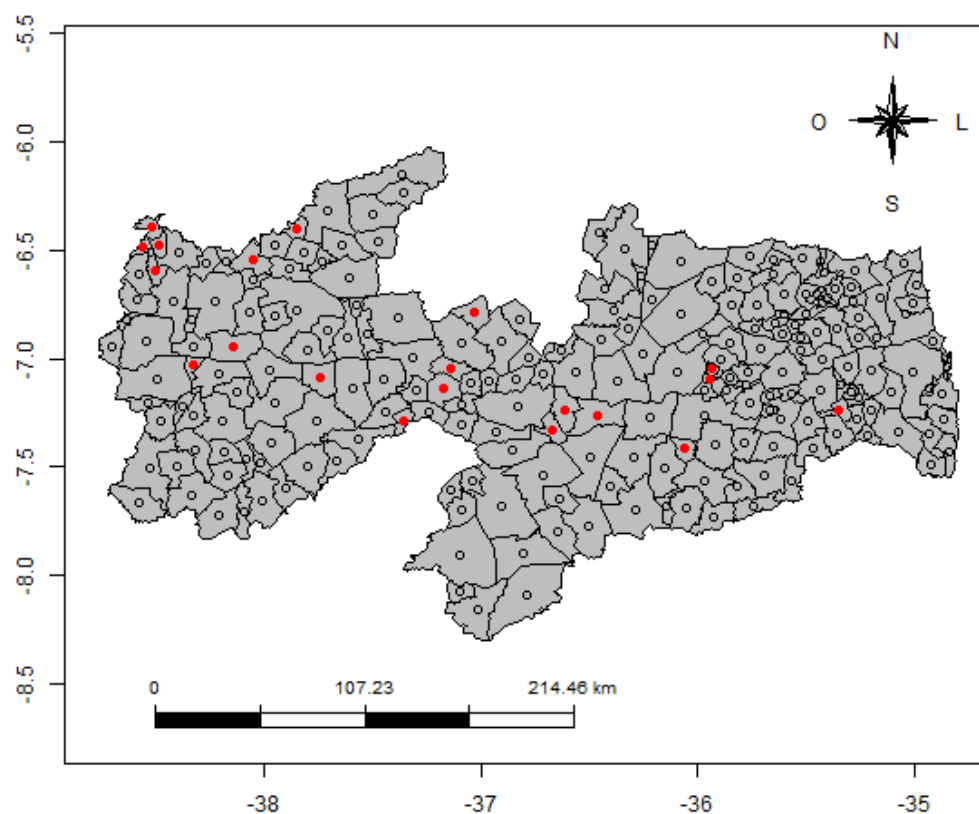
Figura 14 - Mapa dos valores- p do método Getis Ord dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte : Elaborado pelo próprio autor, 2017.

No mapa do método Besag Newell para o dengue em 2011(Fig. 15), observou-se diversos conglomerados espaciais em todas as regiões. Observou-se uma concentração dos conglomerados na mesorregião do Sertão Paraibano, já a mesorregião da Mata Paraibana apresentou o menor número de conglomerados espaciais.

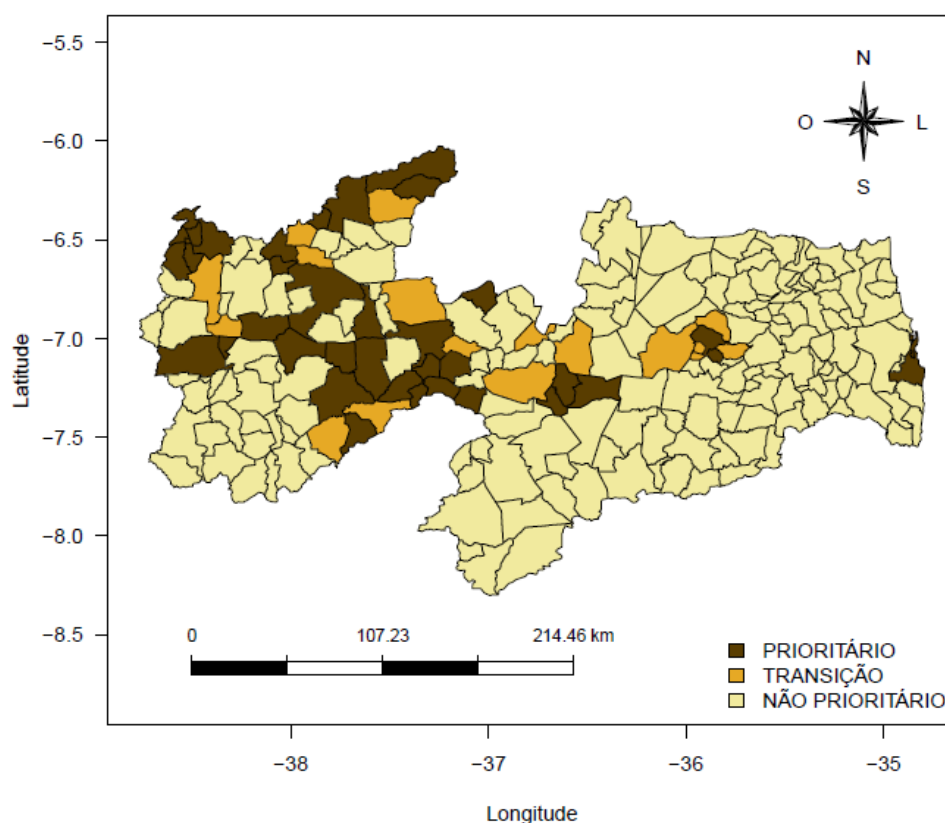
Figura 15 - Mapa dos valores- p do método Besag Newell dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte : Elaborado pelo próprio autor, 2017.

A combinação dos três métodos de agrupamento espacial (estatística Scan espacial, método Besag Newell e estatística Getis-Ord), através da aplicação da arquitetura que utilizou como combinador a votação média ponderada (Figura 10), gerou-se um mapa que está representado na Figura 16, onde podem ser identificados 37 municípios prioritários, 17 municípios em transição (podem se tornar prioritários ou deixar de sê-lo) e 169 não prioritários. Comparando os mapas resultantes dos três métodos (Figuras 13-15) com o mapa da combinação dos métodos na Figura 16 percebe-se a coincidência na concentração dos municípios do sertão e agreste paraibano, como também em cidades de grande contingente populacional.

Figura 16 - Mapa de decisão da combinação dos métodos de aglomeração espacial dos casos de Dengue. Paraíba, Brasil, 2011, através da votação média ponderada



Fonte : Elaborado pelo próprio autor, 2017.

6.2 ESTUDO DE CASO II

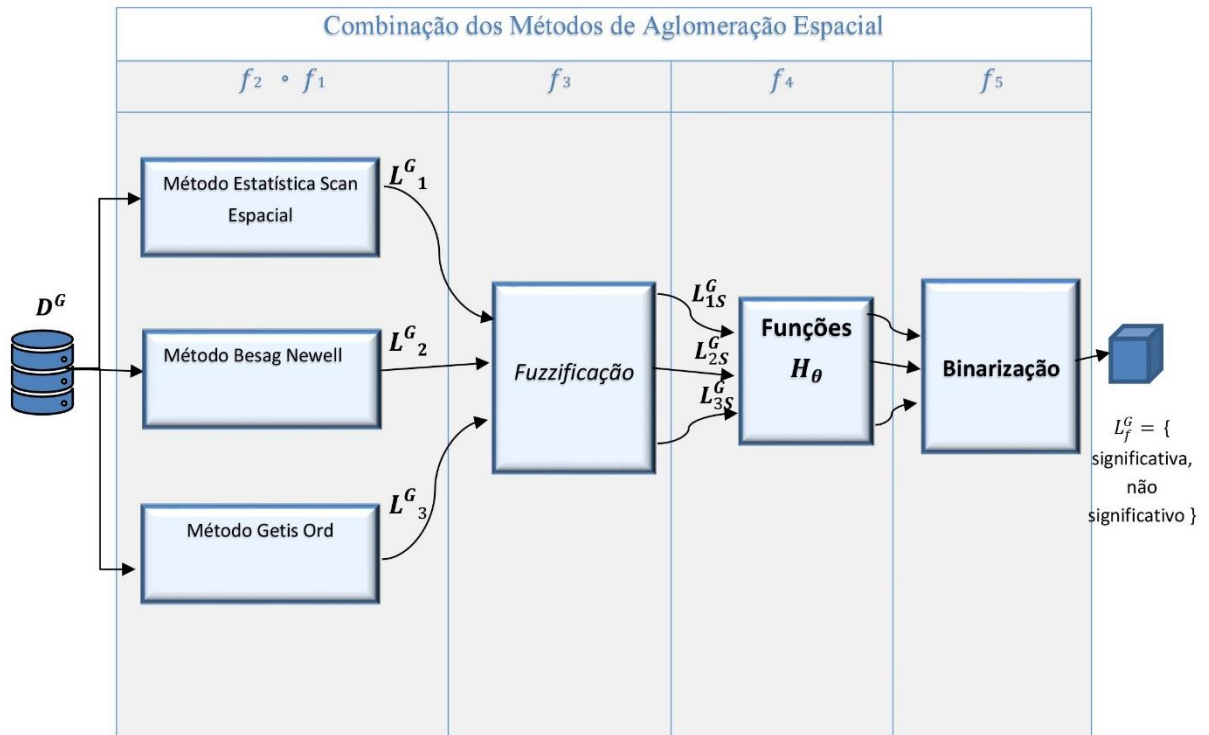
No segundo estudo de caso, foram utilizados os mesmos dados e métodos de aglomeração espacial (Figura 13-15) do primeiro caso, tornando possível a comparação dos resultados. O mapa da RIE (Figura 12) foi utilizado também como referência para análise dos mapas provenientes do método de agregação baseado nas funções H_θ , obtendo os resultados da aplicação da nova arquitetura para combinação dos métodos de aglomeração espacial utilizando como combinador as funções de agregação H_{Max} , H_{Min} , H_{Med} , H_{Arit} e H_{cOWA} (Funções H_θ) para os dados do Dengue, Paraíba, Brasil (Figura 17).

A aplicação da arquitetura inicia com a composição de funções, que pode ser denotada como $f_2 \circ f_1$. Os métodos escolhidos usaram como entrada um banco de dados georreferenciados, ou seja, a função $f_1: D^G \rightarrow \Omega^G$, onde G são os municípios (geo-objetos), D são os números de casos do Dengue e a população de cada município (geo-objeto) e onde $\Omega \in [0,1]$, é o p -valor calculado pelo método de aglomeração espacial para cada geo-objeto. A segunda função é um mapeamento $f_2: \Omega^G \rightarrow L^G$, que fornece um mapa cujo conjunto de rótulos de classe é L^G , em que $L^G = \{s_1, \dots, s_k\}$ para cada objeto geográfico em G , nesse caso o nível de informação fornecido é tipo 3.

A função f_3 é realizada quando há necessidade de padronização da informação, embora opcional nesse caso, foi operacionalizada com intuito de realizar *fuzzificação* da informação. Então nesse caso, podemos adotar uma função de *fuzzificação* em f_3 , gerando informação de saída do Tipo 3 para todos os métodos de aglomeração espacial presentes no conjunto, neste caso, $f_3: L^G \rightarrow L_S^G$, onde $L_S^G = \{(1 - S_1), (1 - S_2), \dots, (1 - S_k)\}$, ou seja, os graus de pertinência intervalados em $[0,1]$.

A quarta função f_4 consiste na aplicação do combinador, neste caso foram as funções de agregação H_{Max} , H_{Min} , H_{Med} , H_{Arit} e H_{cOWA} (Funções H_θ) partindo da combinação dos graus de pertinência. Nesse caso, a combinação foi realizada individualmente para cada função H . Então, a função f_4 do conjunto L_S^G , é dada por $f_4: [0,1]^n \rightarrow [0,1]$, $n=3$. Então, neste estudo caso optou-se por adicionar após a função f_4 uma função de binarização f_5 , com a finalidade de comparar com os resultados de Holmes (2015). Então, $f_5: [0,1] \rightarrow L_f^G$ para o mapa de decisão final, tem o domínio $f_5: [0,1]$ e seu co-domínio definido como $L_f^G = \{\text{significativo}, \text{não significativo}\}$. Por fim, foram gerados 5 mapas de suporte a decisão referente a combinação de cada função H_{Max} , H_{Min} , H_{Med} , H_{Arit} e H_{cOWA} .

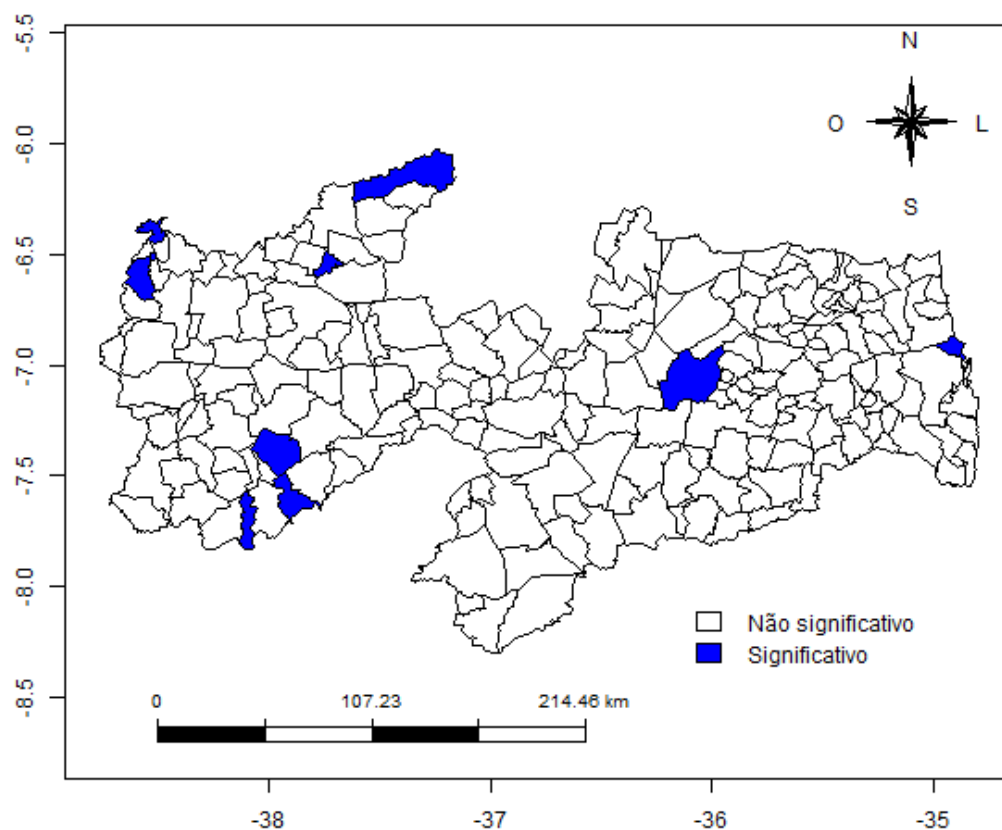
Figura 17 – Aplicação da nova arquitetura para Combinação dos Métodos de Aglomeração Espacial utilizando como combinador as funções de agregação H_{Max} , H_{Min} , H_{Med} , H_{Arit} e H_{COWA} (Funções H_θ) para os dados do Dengue, Paraíba, Brasil



Fonte: Elaborada pelo próprio autor, 2019.

O mapa de decisão resultante da agregação dos métodos utilizando a função H_{Min} (Figura 18), para o ano de 2011, rotulou 11 municípios (geo-objetos) significativos (valor- $p < 0,05$), também chamados de aglomerados espaciais, espera-se uma alta coincidência com as regiões de alto risco apresentadas no mapa do RIE (Figura 13), mas só houveram 03 municípios de alto risco detectados através deste método de agregação, detectando em sua maioria regiões de valores de baixos risco, não sendo considerado um método tão vantajoso.

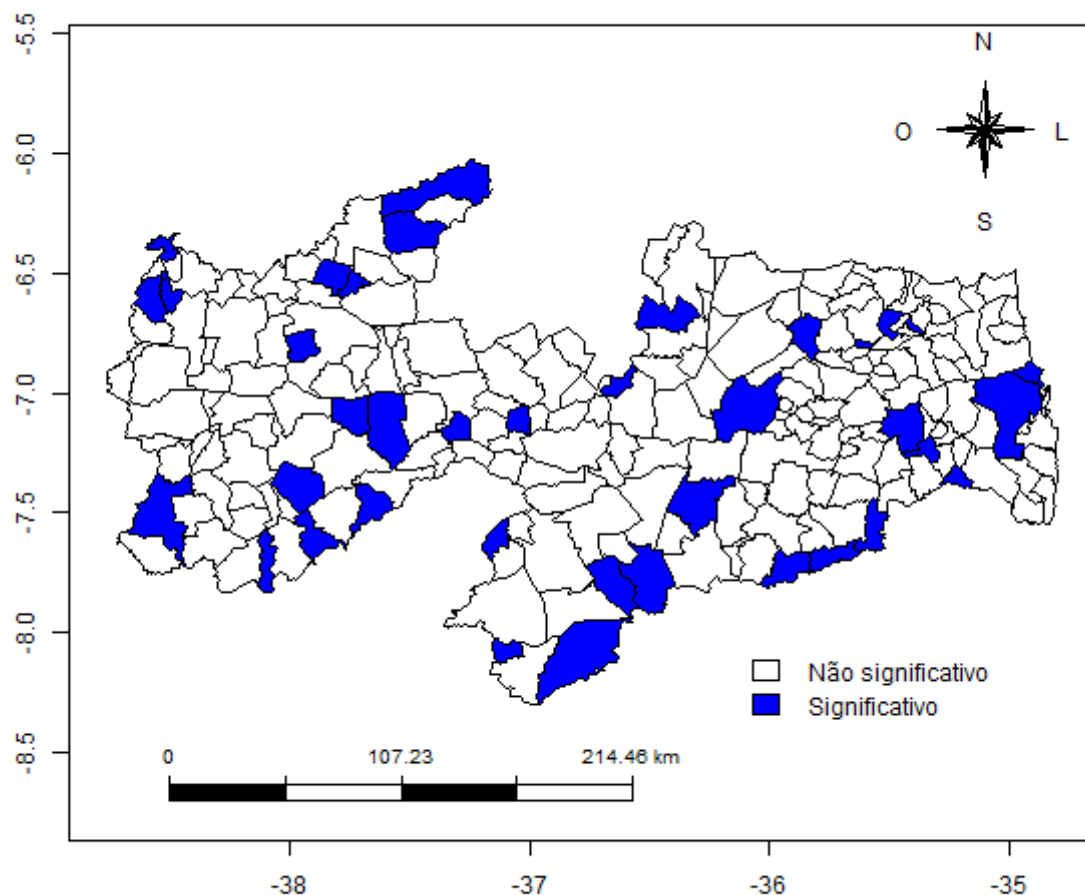
Figura 18 - Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Min} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte: Elaborado pelo próprio autor, 2018.

Em relação ao mapa de decisão resultante da agregação através da função H_{Max} (Figura 19), foram identificados 37 municípios significativos. Comparando com o Mapa do RIE (Figura 12), percebeu-se que o mapa de agregação do Máximo (Figura 19) coincide com as regiões de valores de alto risco, embora também tenha detectado regiões de valores de baixo risco.

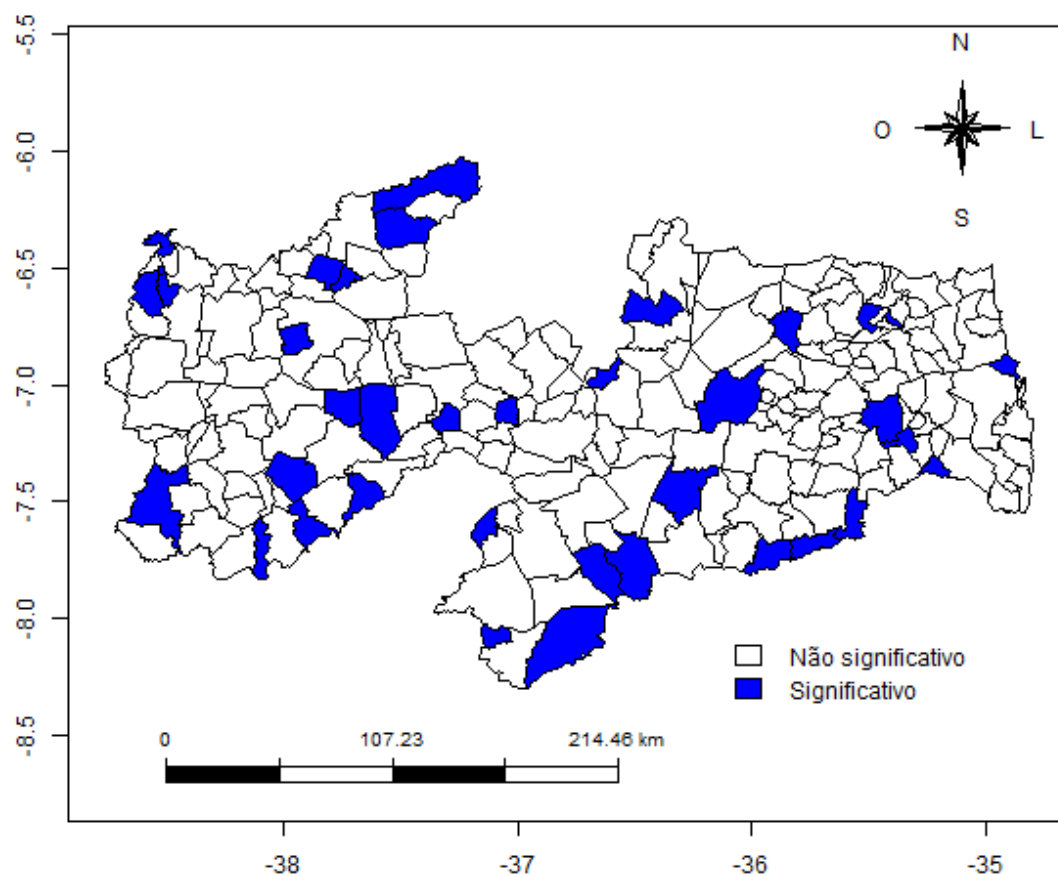
Figura 19 - Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Max} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte: Elaborado pelo próprio autor, 2018.

Ao realizar a agregação dos métodos de aglomeração espacial através da função H_{Medi} , foi gerado um mapa de decisão (Figura 20), tendo que para o ano de 2011, foram identificados 36 municípios (geo-objetos) significativos (valor- $p < 0,05$), também chamados de aglomerados espaciais, este resultado é similar a agregação do máximo (Figura 19), mas detecta 01 município a menos de baixo risco (Figura 12), tornando-o mais vantajoso pois espera-se que sejam detectados os geo-objetos de alto risco.

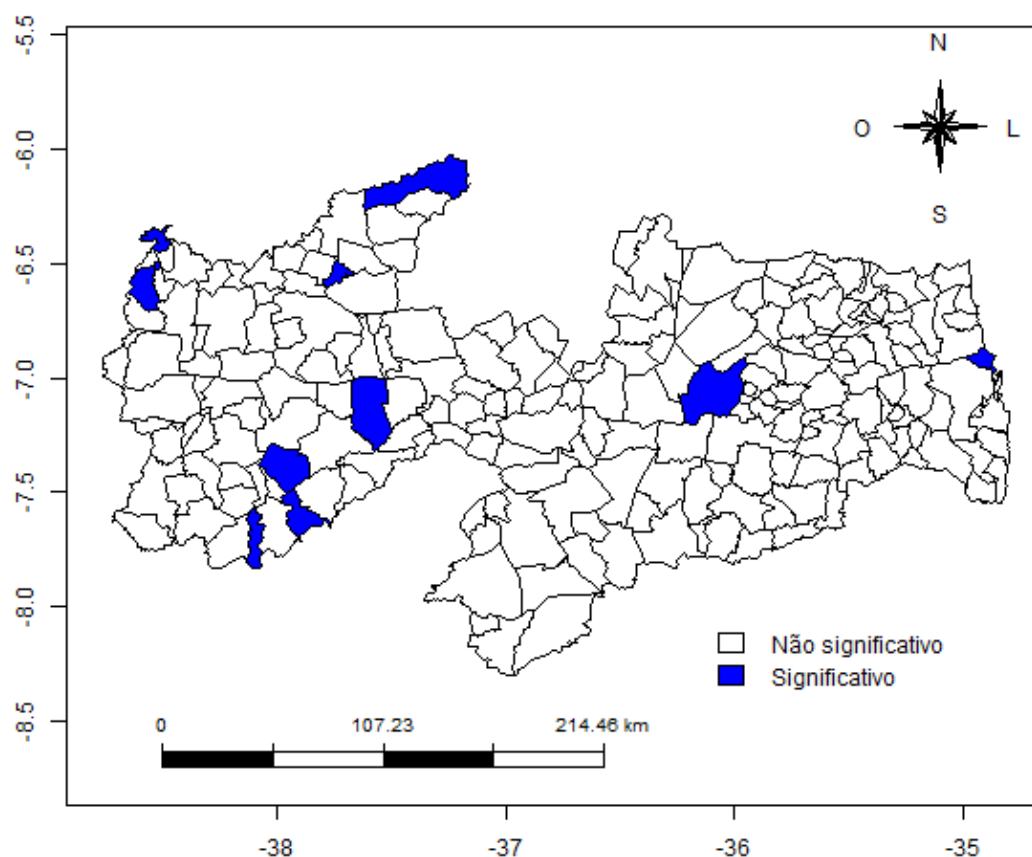
Figura 20 - Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Med} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte: Elaborado pelo próprio autor, 2018.

Em relação ao mapa de decisão gerado a partir da agregação através da função H_{Arit} (Figura 21), foram identificados 12 municípios significativos. Comparando com o Mapa do RIE (Figura 12), percebeu-se que o mapa de agregação do H_{Med} (Figura 21) coincide com as regiões de valores de baixo risco, detectando um município de alto risco.

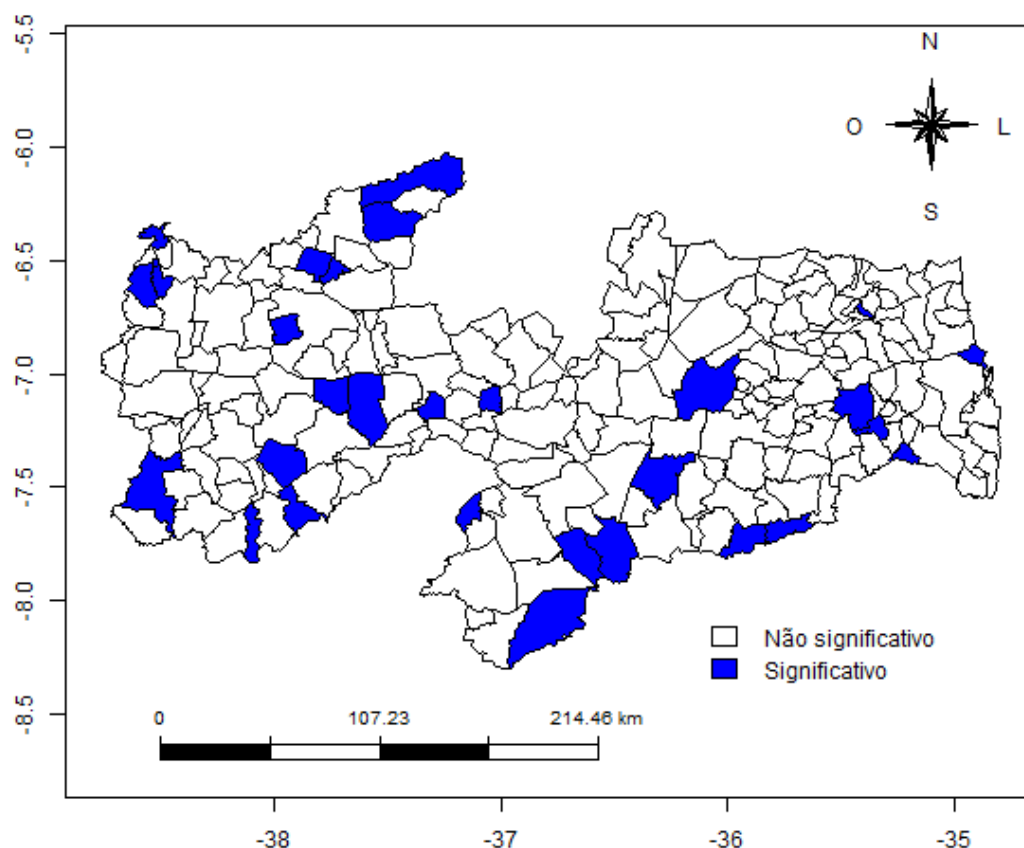
Figura 21 - Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{Arit} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte: Elaborado pelo próprio autor, 2018

Através da análise do mapa de decisão gerado pela função de agregação H_{COWA} (Figura 22), para o ano de 2011, foram identificados 30 municípios (geo-objetos) significativos (valor- $p < 0,05$), também chamados de aglomerados espaciais. Comparando com o Mapa do RIE (Figura 13), percebeu-se que os municípios detectados no mapa de decisão (Figura 22) coincidem com as regiões de valores de alto risco apresentando uma distribuição heterogênea em todo estado.

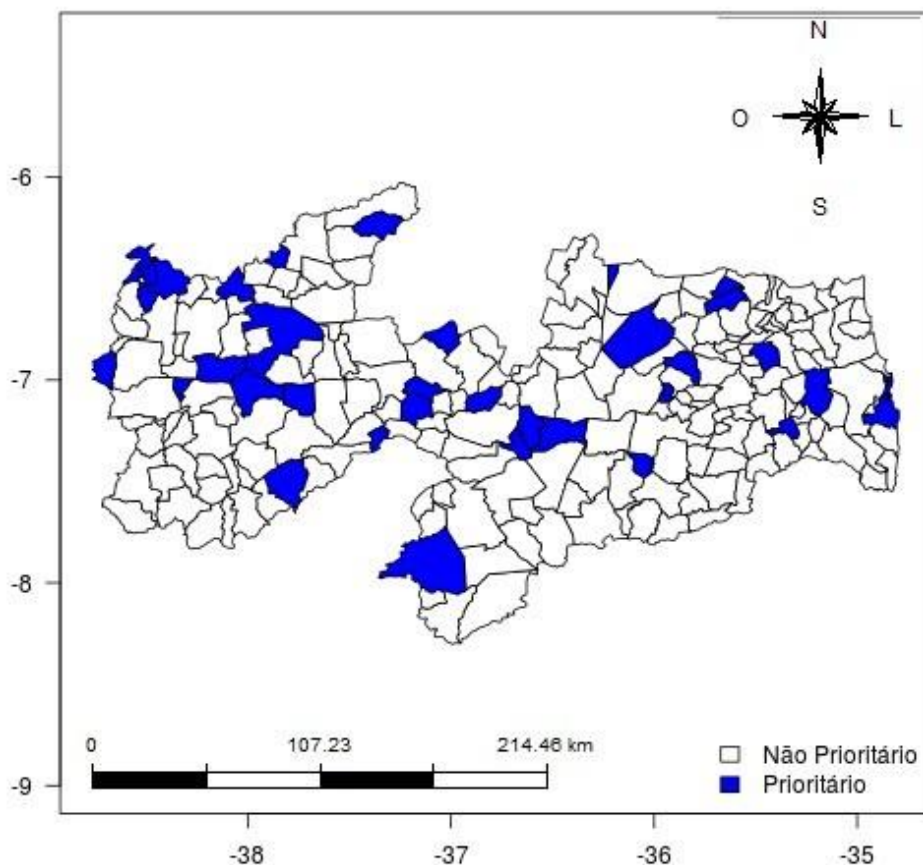
Figura 22 - Mapa de decisão da combinação dos métodos de aglomeração espacial através da função H_{COWA} dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2011



Fonte: Elaborado pelo próprio autor, 2018.

Ao analisar os cinco métodos de agregação utilizados para este estudo, tem-se que os que apresentaram melhores resultados foram a função de agregação H_{Med} (Figura 20), seguido pela H_{Max} (Figura 19) e H_{COWA} (Figura 22). Embora estes tenham apresentado resultados similares o método de agregação da mediana foi o método que mais detectou regiões de alto risco, sendo considerado o melhor resultado, utilizando-se o comparativo com o mapa de RIE (Figura 12). Quando comparamos todos os mapas resultantes das funções H (Figura 18-22) com o mapa da combinação dos métodos de aglomeração espacial através da votação por maioria da arquitetura proposta por Holmes (2015) (Figura 23), percebe-se que há uma similaridade com os resultados da H_{Med} (Figura 20), seguido pela H_{Max} (Figura 19) e H_{COWA} (Figura 22).

Figura 23 - Mapa de decisão da combinação dos métodos de aglomeração espacial através da votação por maioria da arquitetura proposta por Holmes(2015) para o ano de 2011



Fonte: Holmes, 2015.

6.3 ESTUDO DE CASO III

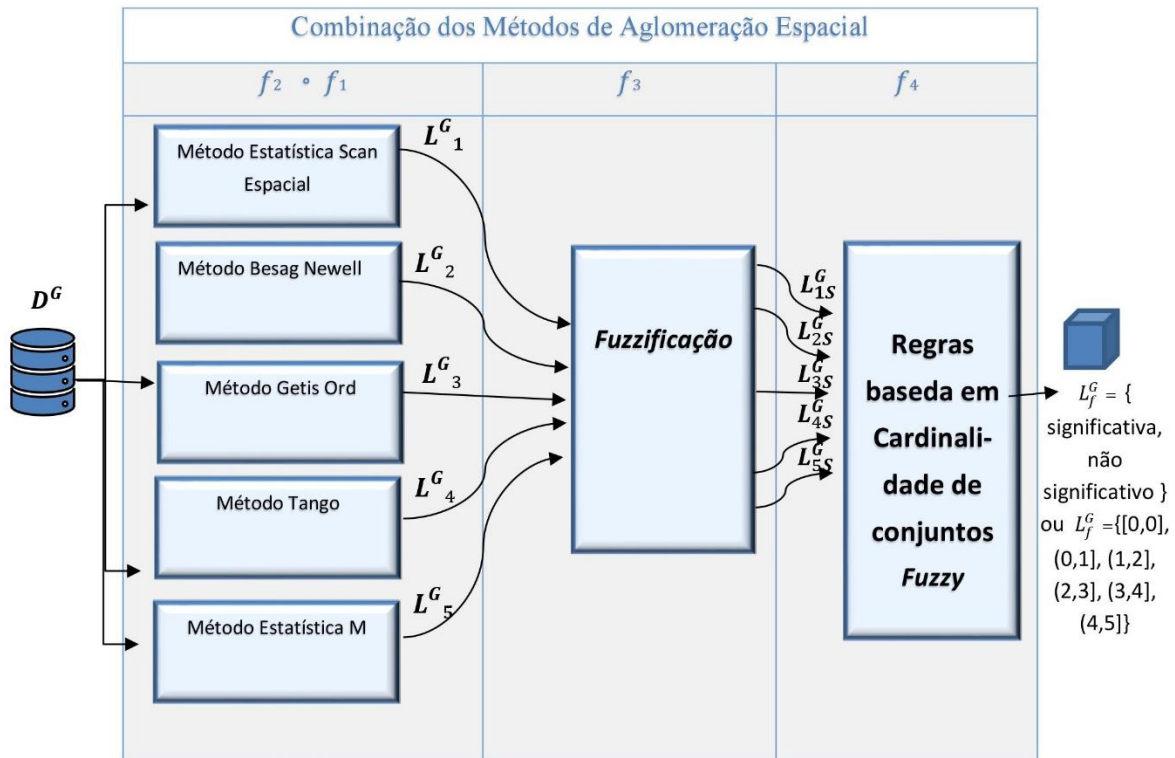
O terceiro estudo de caso utilizou dados do Dengue mais recentes do ano 2017, no estado da Paraíba, Brasil. Para análise espacial foram escolhidos cinco métodos de aglomeração espacial: Estatística Scan espacial, Método Besag Newell, Getis Ord, Estatística M e Tango. A nova arquitetura foi aplicada utilizando votação por maioria *fuzzy* utilizando cardinalidade de conjuntos *fuzzy* e o *plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy* que foram propostas neste trabalho, também foi utilizado a cardinalidade de conjuntos *fuzzy* para construção do mapa por classes (Figura 24). Ao fim, os dados foram aplicados na arquitetura anterior proposta por Holmes (2015) (Figura 6) com a finalidade de comparação.

A aplicação da arquitetura inicia com a composição de funções, que pode ser denotada como $f_2 \circ f_1$. Os cinco métodos escolhidos usaram como entrada um banco de dados georreferenciados, ou seja, a função $f_1 : D^G \rightarrow \Omega^G$, onde G são os municípios (geo-objetos), D são os números de casos do Dengue e a população de cada município (geo-objeto) e onde $\Omega \in [0,1]$, é o p -valor calculado pelo método de aglomeração espacial para cada geo-objeto. A segunda função é um mapeamento $f_2 : \Omega^G \rightarrow L^G$, que fornece um mapa cujo conjunto de rótulos de classe é L^G , em que $L^G = \{s_1, \dots, s_k\}$ para cada objeto geográfico em G , nesse caso o nível de informação fornecido é tipo 3.

A função f_3 é realizada quando há necessidade de padronização da informação, embora opcional nesse caso, foi operacionalizada com intuito de realizar *fuzzificação* da informação. Então, podemos adotar uma função de *fuzzificação* em f_3 , gerando informação de saída do Tipo 3 para todos os métodos de aglomeração espacial presentes no conjunto, neste caso, $f_3 : L^G \rightarrow L_S^G$, onde $L_S^G = \{(1 - S_1), (1 - S_2), \dots, (1 - S_k)\}$, ou seja, os graus de pertinência intervalados em $[0,1]$.

A quarta função f_4 consiste na aplicação do combinador, neste caso as regras de combinação foram: votação por maioria *fuzzy* utilizando cardinalidade de conjuntos *fuzzy*, *plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy*, e por fim a cardinalidade de conjuntos *fuzzy*. Estas regras realizaram a combinação dos graus de pertinência, após *fuzzificação*. Nesse caso, a combinação foi realizada individualmente para cada função regra de combinação. Então, a função f_4 do conjunto L_S^G para o mapa de decisão final, tem o domínio $f_4 : \{0,1\}$ e seu co-domínio $L_f^G = \{\text{significativo}, \text{não significativo}\}$. Como também, se domínio f_4 é $[0,1]$, seu co-domínio como $L_f^G = \{[0,0], (0,1], (1,2], \dots, (t-1, t]\}$, t pertence $\mathbb{R}^+$ e é limitante superior do $\Sigma Count$ da votação gerada pelo conjunto de métodos na combinação. Por fim, foram gerados 3 mapas de suporte a decisão referente a combinação de cada regra de votação.

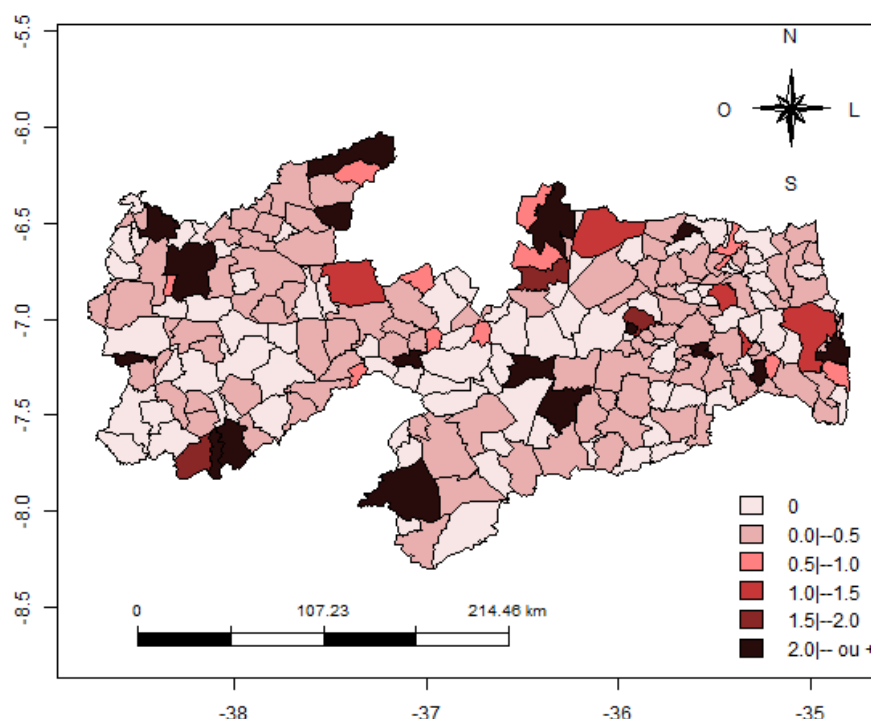
Figura 24 – Aplicação da nova arquitetura para Combinação dos Métodos de Aglomeração Espacial utilizando regras de combinação baseadas em cardinalidade de conjuntos *fuzzy* para os dados do Dengue, Paraíba, Brasil



Fonte: Elaborada pelo próprio autor, 2019.

Foram utilizados os dados de notificações do Dengue no estado da Paraíba, Brasil, para o ano de 2017, que totalizaram 15.738 casos do Dengue no estado (Sinan, 2018). Inicialmente, foi gerado o mapa do RIE (Figura 25) para o referido ano com o intuito de ser utilizado como referência para comparação com os mapas dos métodos de aglomeração espacial e apresentando uma heterogeneidade a distribuição da doença em todo estado.

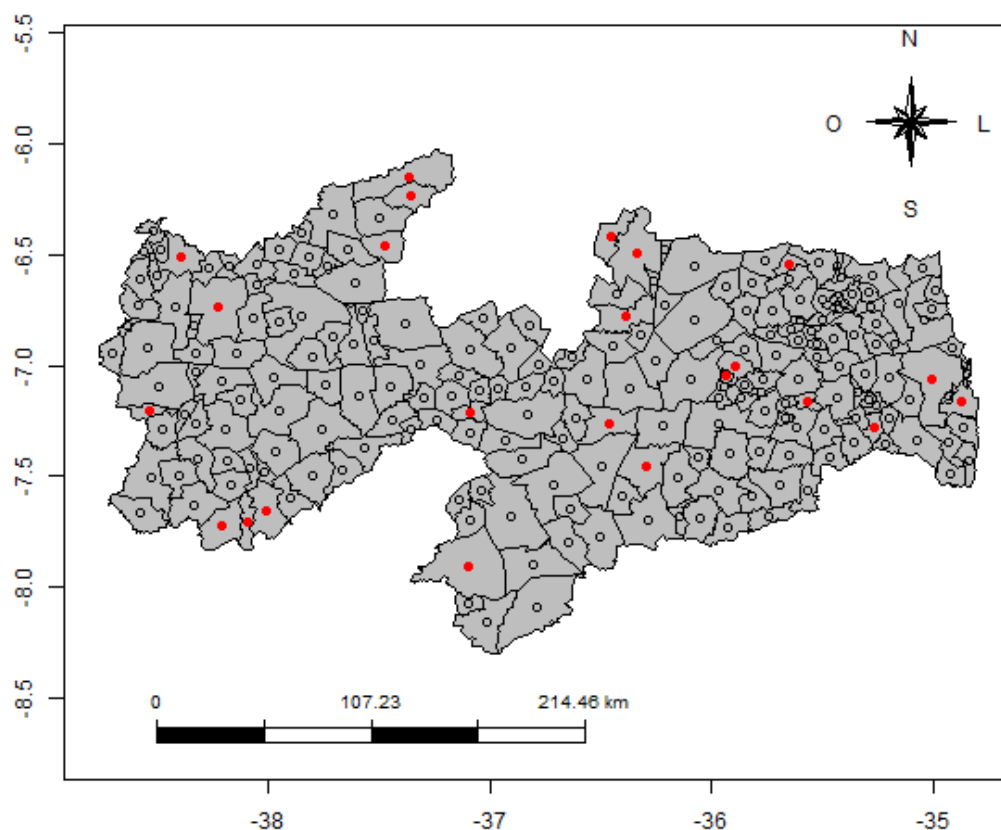
Figura 25- Mapa da Razão de Incidência espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

No mapa da estatística *Scan* Espacial (Figura 26) foram identificados vinte e três (23) municípios (geo-objetos) pertencentes a conglomerados espaciais para o ano 2017, com percentual da população de 0,1%. Comparando-se o mapa do *Scan* espacial com o mapa de RIE (Figura 25), pode-se verificar a coincidência das regiões que apresentam os maiores índices de RIE com as que possuem conglomerados espaciais significativos.

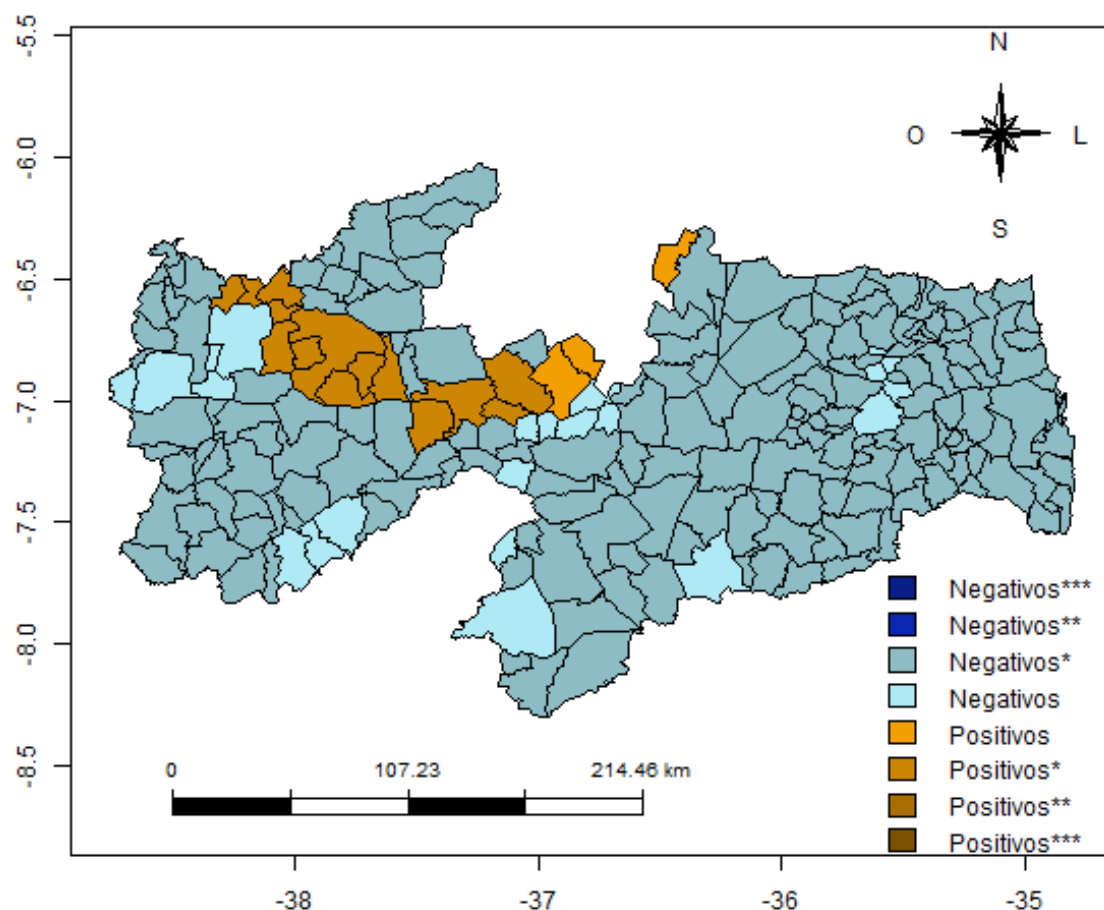
Figura 26 - Mapa da estatística *Scan* espacial dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

No mapa resultante da aplicação do método Getis Ord (Figura 27), foi possível identificar aglomerados espaciais positivos e negativos. Apenas alguns municípios apresentaram aglomerados espaciais de valores altos, que se localizaram em sua maioria na mesorregião do Sertão Paraibano e Borborema. Os aglomerados espaciais de valores baixos estavam distribuídos de forma heterogênea em toda a região da Paraíba.

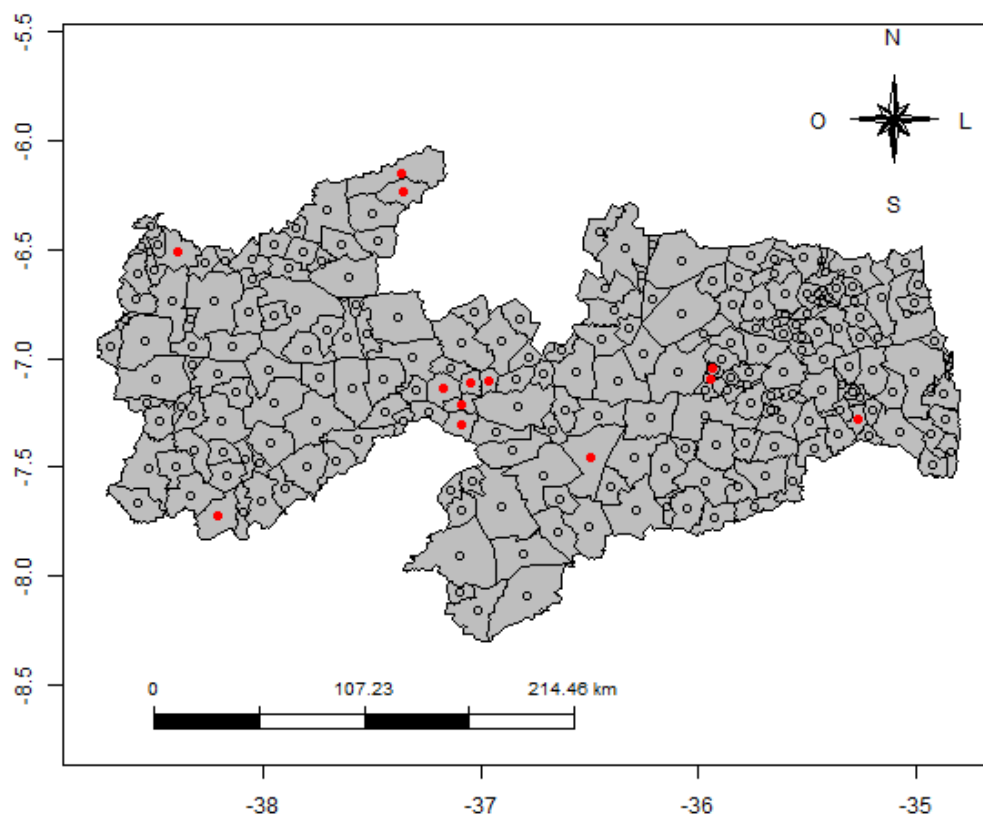
Figura 27 - Mapa do método Getis Ord dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

Em relação ao mapa do método Besag Newell para o dengue em 2017 (Figura 28), observou-se diversos conglomerados espaciais em todas as regiões. Observou-se uma concentração dos conglomerados na mesorregião do Sertão Paraibano e Borborema, já a mesorregião da Mata Paraibana apresentou o menor número de conglomerados espaciais.

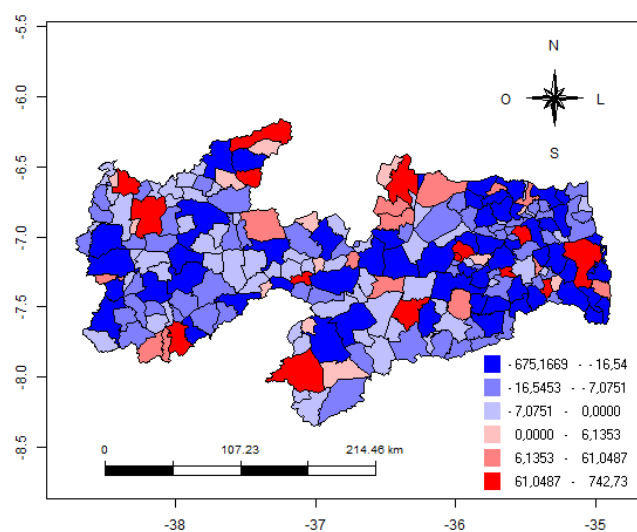
Figura 28 - Mapa do método Besag Newell dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



Fonte : Elaborado pelo próprio autor.

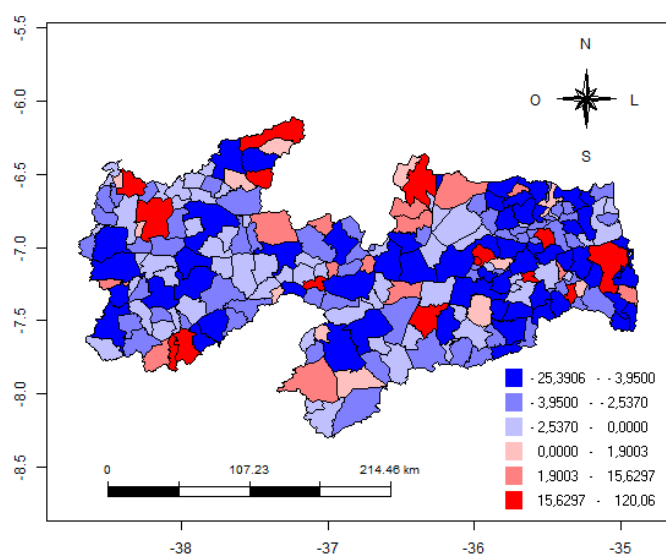
Os mapas gerados pela Estatística M e Tango apresentaram resultados similares e a cor azul escuro indica a presença de conglomerados espaciais de valores baixos ou nulos e a cor vermelha indica a presença de conglomerados espaciais de valores altos. Ao compará-los com o mapa de RIE (Figura 25), observou-se que a maioria dos aglomerados espaciais detectados corresponderam aos municípios com valores altos da razão de incidência espacial, conforme as figuras 29-30.

Figura 29 - Mapa do método Tango dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

Figura 30 - Mapa do método Estatística M dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



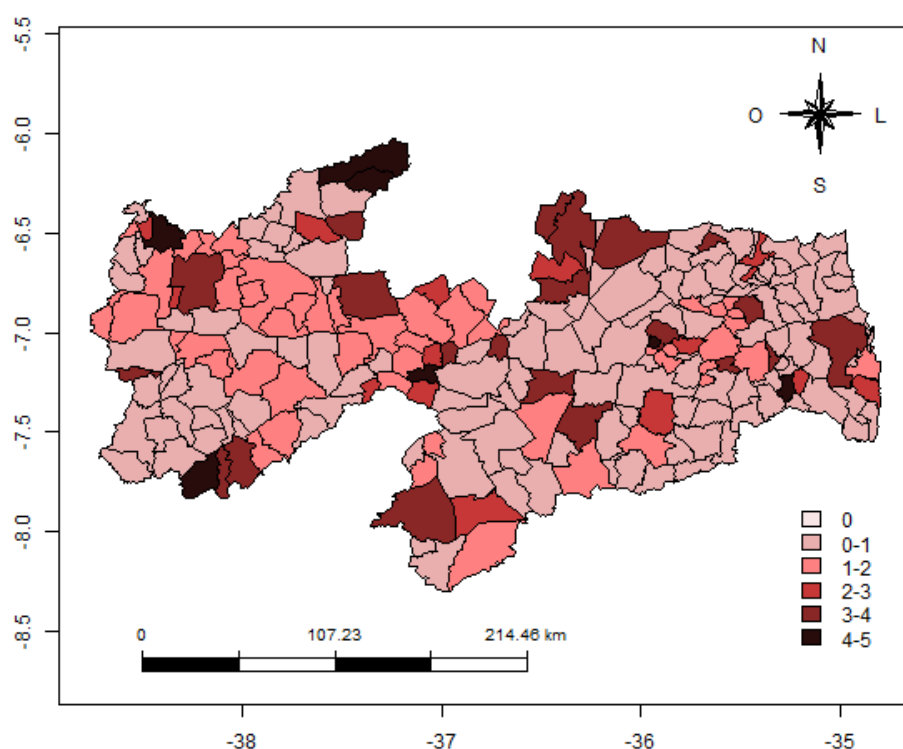
Fonte: Elaborado pelo próprio autor, 2018.

Os métodos de aglomeração espacial usados como entrada para combinação neste estudo foram dispostos nas figuras 26-30, quando se compara os resultados dos métodos, percebe-se que apresentam diferenças em relação aos municípios considerados significativos.

Baseado na diferença existente entre os resultados dos métodos de aglomeração espacial foi realizada a combinação destes através da arquitetura mostrada na figura 24.

Visualiza-se, na figura 31 o mapa de votação dos métodos de aglomeração espacial que foi construído a partir do somatório dos graus de pertinência derivados dos métodos de aglomeração espacial, tornando possível agrupar os municípios por faixas de significância, para isto utilizou como agregador o Σ count.

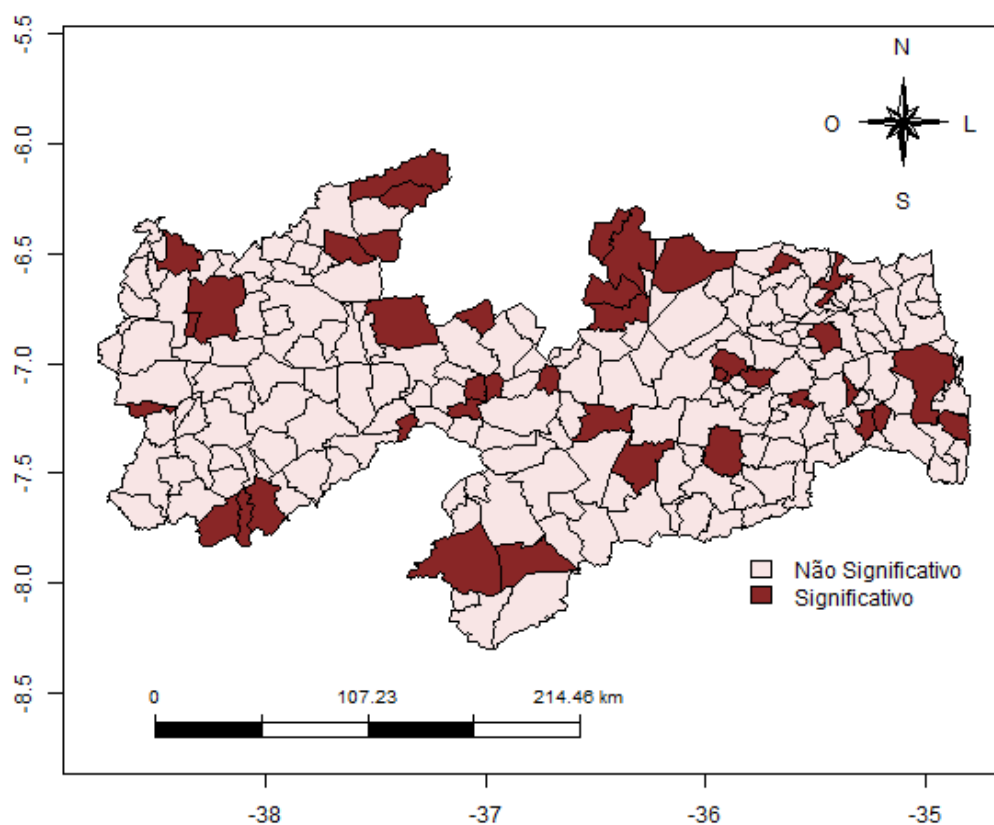
Figura 31 - Mapa de decisão dos métodos de aglomeração espacial utilizando a cardinalidade de conjuntos *fuzzy*. Paraíba, Brazil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

O mapa observado na figura 32 é resultante do somatório das funções de pertinência dos métodos de aglomeração espacial, através da votação por maioria *fuzzy* que utiliza cardinalidade de conjuntos *fuzzy*. Foram considerados significativos os municípios que estavam acima do ponto de corte estabelecido, este método preserva a informação, que é perdida com a binarização, permitindo uma abrangência de municípios considerados significativos.

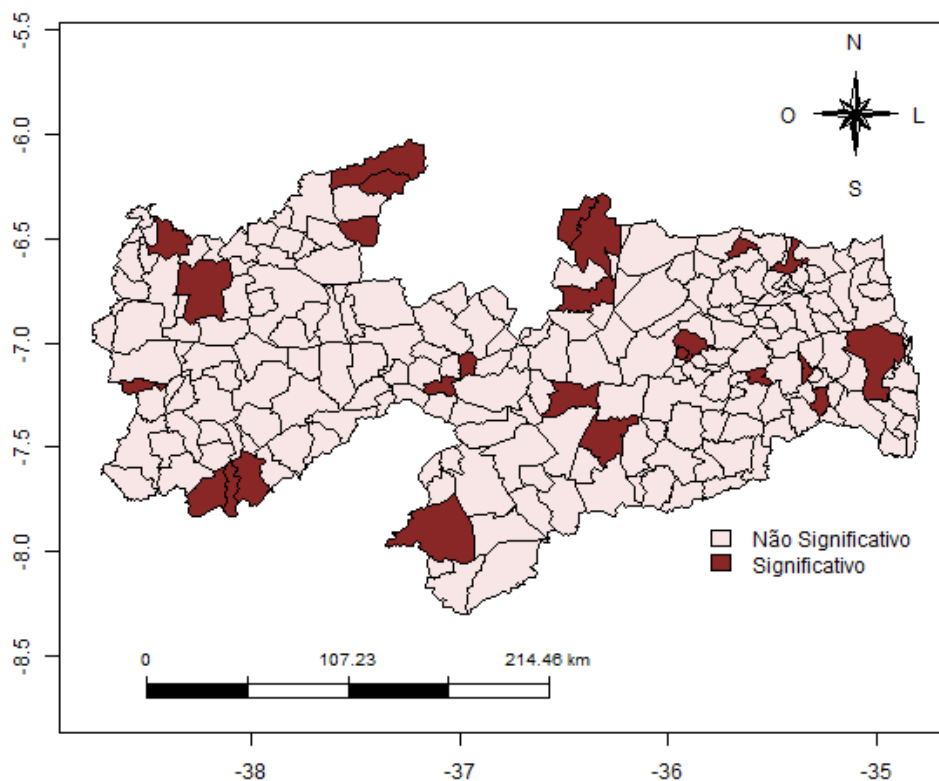
Figura 32 - Mapa de decisão da votação por maioria *fuzzy* utilizando cardinalidade de conjuntos *fuzzy* dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

O mapa observado na figura 33 é resultante do somatório das funções de pertinência dos métodos de aglomeração espacial utilizando *Plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy*, sendo considerado significativo os municípios que apresentassem o somatório das funções de pertinência a favor maior que os da classe contra. Comparando os mapas (fig. 30-31), quando utilizada a *Plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy* o resultado apresentado é melhor que a votação por maioria *fuzzy* utilizando cardinalidade de conjuntos *fuzzy* devido a coincidência maior com municípios de alto risco no RIE (Figura 25). A principal vantagem da utilização do sigma count foi a possibilidade de generalizar o método plurality voting e votação por maioria, já que com a não binarização da informação, os resultados tornam-se diferentes na aplicação destes métodos.

Figura 33 - Mapa de decisão da *Plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy* dos casos de Dengue segundo Município de residência. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

6.3.1 Comparação com arquitetura existente de combinação de métodos de aglomeração espacial (HOLMES, 2015)

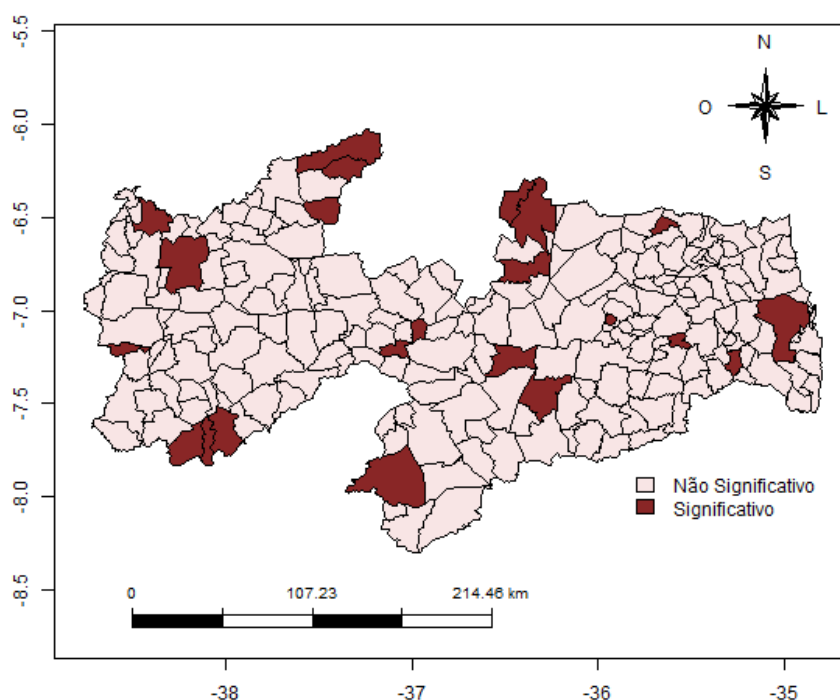
Para comparação com os resultados obtidos da combinação a partir das votações baseadas na cardinalidade *fuzzy* apresentados nas Figuras 31-33, foi replicado o modelo encontrado na literatura que realiza a combinação de métodos de aglomeração espacial para o ano de 2017 (HOLMES, 2015; HOLMES, et al, 2018). Os métodos de aglomeração espacial usados como entrada estão dispostos nas figuras 26-30.

Observa-se na figura 34 o mapa resultante da combinação dos métodos de aglomeração espacial através da votação por maioria, que surge a partir da utilização de uma informação binarizada, classificando os municípios que obtiveram 3 votos ou mais como significativos. Visualiza-se-se na figura 35 o mapa resultante da combinação dos métodos de aglomeração espacial através do plurality voting, que também utiliza uma informação

binarizada, tornando o resultado igual ao da votação por maioria simples (Figura 34), ou seja, o método utilizado por Holmes (2015), não permite diferença nos resultados, pois binariza a informação, classificando também os municípios que obtiveram 3 votos ou mais como significativos.

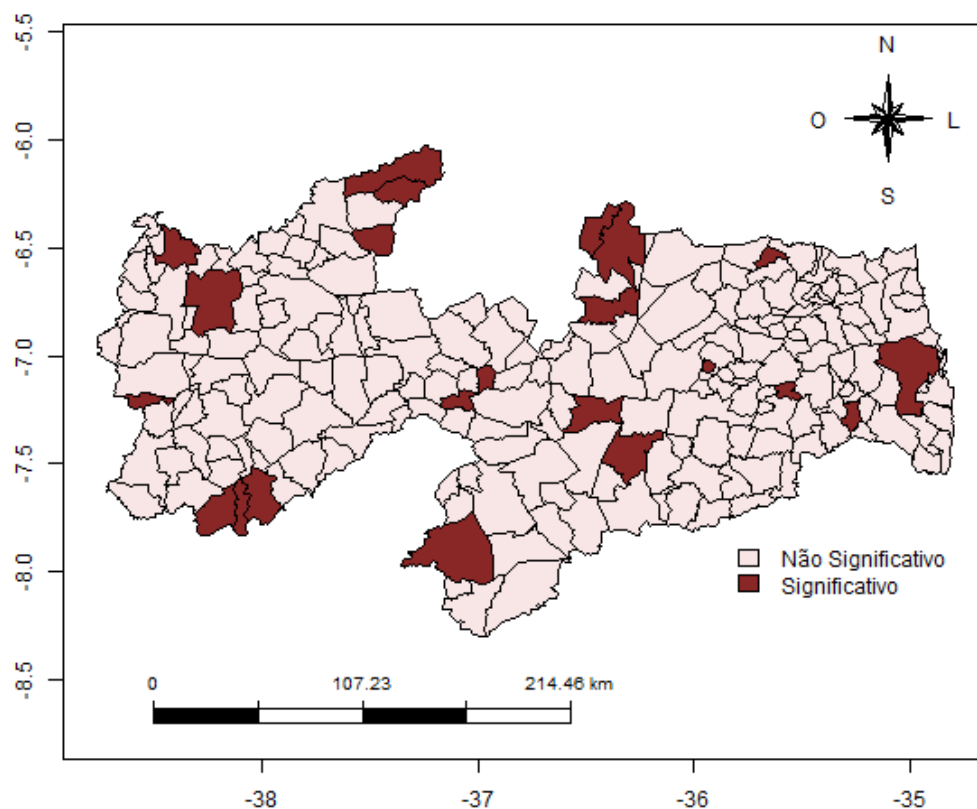
O novo modelo proposto que utiliza cardinalidade de conjuntos *fuzzy* (Figura 31-33), demonstra vantagem em relação ao método anterior (Figura 34-35), pois ao utilizar o *plurality voting fuzzy*, torna-se menos restritivo e preserva a informação não utilizada anteriormente devido a binarização da informação.

Figura 34 - Mapa de decisão da combinação dos métodos de aglomeração espacial utilizando a votação por maioria. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

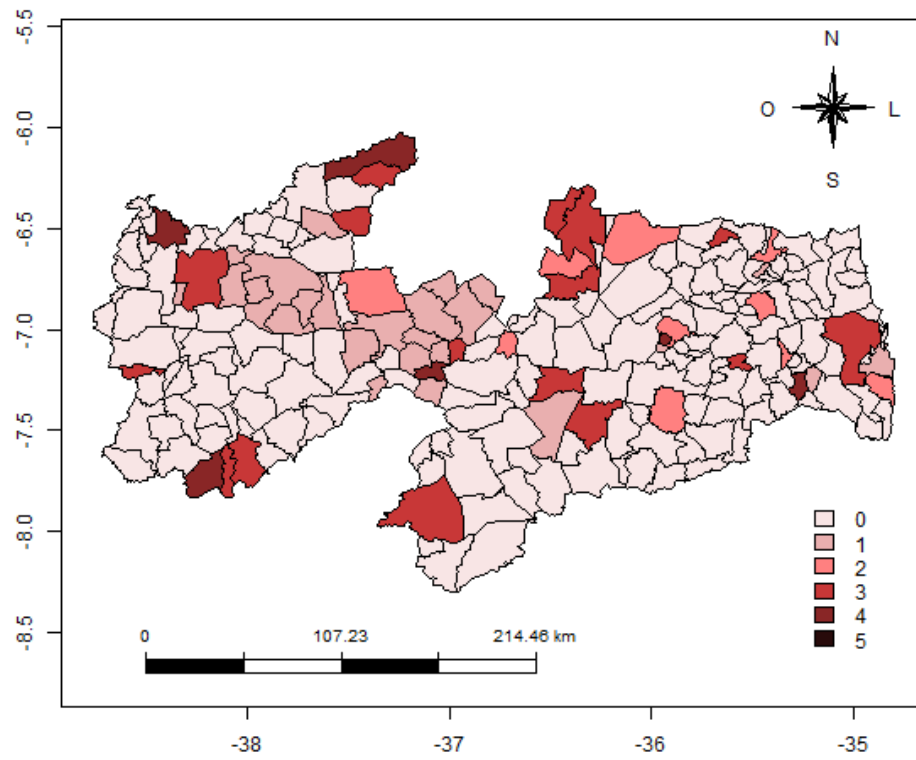
Figura 35 - Mapa de decisão da combinação dos métodos de aglomeração espacial utilizando plurality voting. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor.

Na figura 36, tem-se o mapa de votos resultante do método utilizado por Holmes (2015; 2018), comparado com o mapa derivado do novo método (Figura 31) percebe-se que ao utilizar os graus de pertinência divididos em classes, tem-se a informação preservada, dando uma flexibilidade maior frente a tomada de decisão.

Figura 36 - Mapa de votação da combinação dos métodos de aglomeração espacial. Paraíba, Brasil, 2017



Fonte: Elaborado pelo próprio autor, 2018.

“Aprendendo irás ensinar e ensinando irás aprender.”

Phil Collins

8 DISCUSSÃO

Como referido anteriormente o dengue é uma doença acarretada por vírus e transmitida por vetor, aonde sua incidência tem aumentado de forma exacerbada, representando um problema de saúde pública o qual pode afetar qualquer pessoa independentemente de fatores socioeconômico e culturais. Vale ressaltar, que a distribuição geográfica dos sorotipos da dengue varia imprevisivelmente ao longo do tempo e do espaço, provocando risco contínuo de infecção em populações de áreas endêmicas, apesar das ações de controle do vetor. É possível afirmar que o controle do dengue é tecnicamente viável, desde que haja apoio técnico e financeiro coordenado para programas nacionais (WHO, 2016).

Nesse sentido, Gusman et al. (2016) afirma que a disseminação do vírus da dengue ocorre a nível global, sendo sua transmissão fortemente influenciada pela precipitação, temperatura, urbanização e distribuição do principal mosquito vetor *Aedes aegypti*, levando a uma variação espacial. Os padrões epidemiológicos dessa doença são alarmantes tanto para a saúde quanto para a economia mundial, sendo considerada como uma morbidade do futuro devido a tendências de aumento da urbanização, escassez de água e, possivelmente, mudança ambiental.

Diante do exposto, podemos notar que a complexidade na dinâmica da circulação dos sorotipos dos vírus do dengue, apontam a necessidade de estudos que determinem a sua distribuição espacial. De acordo com Bhatt et al. (2013), o mosquito vetor da doença, o *Aedes aegypti*, vive em ambientes de grande densidade demográfica e um dos fatores principais que contribuiu para acelerar a disseminação da doença foi a urbanização recente principalmente nas regiões tropicais.

Vários estudos sobre dengue no mundo, já incluem as informações georreferenciadas com o intuito de entender como ocorre a disseminação da doença, além de propor medidas preventivas e de combate ao dengue. No estado da Paraíba, Brasil pesquisas vem sendo desenvolvidas nesse sentido, utilizando técnicas que aprimoram a utilização dos métodos de aglomeração espacial (MELO; MORAES, 2018; HOLMES; MORAES; VIANNA, 2015; SÁ, et al. 2015; HOLMES; MORAES; VIANNA, 2018 SÁ, et al. 2018).

Portanto, com base na análise espacial e os mapas gerados nesse estudo para os anos de 2011 e 2017, foi possível observar que há uma heterogeneidade na distribuição geográfica da dengue no Estado da Paraíba, porém há uma maior concentração em áreas com grande densidade populacional e nas mesorregiões do sertão e agreste paraibano dos municípios que

apresentaram os maiores valores de RIE, além de serem classificados como conglomerados estatisticamente significativos.

Resultado semelhante foi obtido no estudo realizado por Melo e Moraes (2018) um sistema espacial de suporte à decisão para gestão do combate ao dengue usando lógica *fuzzy* foi desenvolvido no estado da Paraíba entre os anos de 2009 a 2013 para detectar áreas de prioridade de combate a esse agravo. O estudo realizado por Holmes (2015) obteve também resultados similares apontando que a mesorregião do Sertão Paraibano apresentou as maiores concentrações de aglomerados do dengue no estado e a mesorregião do Litoral/Mata Paraibana as menores concentrações de aglomerados.

Os municípios que foram considerados significativos são propícios à dinâmica de transmissão da doença devido a indisponibilidade de serviços de saneamento ambiental adequados, como também a falta de abastecimento água principalmente no sertão paraibano. Segundo San Pedro, et al. (2009), a oferta precária desses serviços, principalmente quanto ao abastecimento de água, pode levar à prática de estocagem de água em recipientes que de forma inadequada podem tornar-se potenciais locais de reprodução do vetor. Outros fatores apontados que propiciam a transmissão foram: o grande fluxo populacional entre localidades, a alta densidade populacional nas áreas metropolitanas, a urbanização desordenada, responsável pela precariedade das condições sócio sanitárias, bem como a pouca eficácia dos programas governamentais de controle da doença.

Em pesquisa realizada por Fantinati, et al (2013) observou-se que a maior parte da população que possui poucos anos de estudo, apresentam uma predominância de registro maior para o dengue. Vale salientar, os indivíduos mais pobres tendem a procurar mais os serviços de saúde por problemas de doença, em detrimento da prevenção ou da realização de exames de rotina, como observado na população com maior escolaridade. Este quadro está condicionado ao fato de os grupos menos privilegiados terem reduzida oportunidade de acesso, desconhecendo seu estado de saúde (NERI; SOARES, 2002).

É sabido que a erradicação da dengue é incerta, pois depende de vários fatores relacionados à doença. Sendo a medida mais eficaz e economicamente viável, até o momento, é a de combate ao vetor (*A.aegypti*). Nesse sentido, os métodos de aglomeração espacial tornam-se uma ferramenta útil para os estudos epidemiológicos da doença. Os métodos de aglomeração espacial analisam o fenômeno em estudo, identificando os aglomerados significativos e não significativos, que quando utilizados individualmente não refletem exatamente a realidade do fenômeno estudado (SÁ et al., 2018).

Um avanço nas áreas da epidemiologia e estatística espacial foi o desenvolvimento de uma arquitetura para combinação dos métodos de aglomeração espacial adotando regras de combinação, visando tornar a tomada de decisão eficiente, visto que estes métodos apresentam uma variedade metodológica que levam a resultados diferentes não refletindo a realidade do fenômeno (HOLMES; MORAES; VIANNA, 2015; HOLMES; MORAES; VIANNA, 2018). O estudo realizado por Holmes (2015), aponta diversas contribuições frente a combinação dos métodos de aglomeração, sendo possível identificar os geo-objetos (municípios) e suas regiões (mesorregiões) como prioritárias e não prioritárias, além da proposição de um mapa de votação mais detalhado, onde são identificados a quantidade de votos que cada geo-objeto recebe durante o processo de votação.

É possível afirmar que com a combinação entre os métodos percebe-se um ganho de informação que torna possível obter resultados satisfatórios e aproximados da realidade do fenômeno, contudo percebeu-se que há uma limitação pois não inclui as incertezas devido a binarização, como exposto anteriormente.

A nova arquitetura proposta nesse estudo que utiliza regras de combinação *fuzzy* para métodos de aglomeração espacial, surge como uma ferramenta que possibilita incluir todas as incertezas inerentes no processo saúde-doença, visto que toda informação é preservada. Corroborando com esta ideia Bassanezi et al. (2001), mostram a necessidade de se estudar o comportamento dinâmico das epidemias, contudo grande parte dos fenômenos relacionados a saúde e adoecimento possuem um alto grau de subjetividade que não é incorporada nos modelos matemáticos tradicionais, portanto os graus de incertezas passam ser modelados pela Teoria *Fuzzy*, formalizada por Zadeh (1965). A utilização da teoria nos estudos relacionado a epidemiologia, deve-se a necessidade de prever a dinâmica de transmissão da doença, além de auxiliar a tomada de decisões sobre as medidas preventivas que possam evitar a ocorrência das epidemias.

O primeiro estudo de caso deste trabalho ao realizar a combinação dos métodos de aglomeração espacial, utilizando a regra de votação média ponderada, obteve um mapa final com a classificação de cada município segundo “municípios prioritários”, “municípios de transição” (que podem se tornar prioridade ou não) e “não prioritário”. Este método permitiu a visualização da distribuição espacial da dengue em todos os municípios da Paraíba, permitindo identificar municípios vulneráveis, podendo ajudar os gestores nas decisões relativas às características específicas de cada município segundo os níveis de prioridade.

Com a aplicação do segundo estudo de caso usando dados epidemiológicos de dengue do estado da Paraíba, Brasil, no ano de 2011, foi possível identificar os municípios

significativos e não significativos para a dengue, utilizando a classe de funções de H_{Θ} (H_{Max} , H_{Min} , H_{Med} , H_{Arit} e H_{COWA}), fornecendo diferentes resultados quando comparadas com o RIE. Neste caso, o melhor resultado foi obtido usando a função H_{Med} , por apresentar uma maior coincidência nos municípios detectados com os municípios considerados significativos no mapa RIE. As funções H_{Max} e H_{COWA} também obtiveram resultados semelhantes, mas detectaram menos municípios considerados significativos no RIE. Com base nesses resultados, é possível afirmar que a combinação de métodos aglomeração espacial usando essa nova classe de regras foi capaz de fornecer resultados aceitáveis, quando comparado com o mapa RIE usado como referência. Por fim, essa nova classe de regras permite também o uso de diferentes abordagens epidemiológicas, pois detecta aglomerados de alta e / ou baixa incidência. Além disso, pode ser usado para melhorar os resultados de métodos de aglomeração espacial em outras áreas de aplicação.

Comparando os resultados obtidos nos dois estudos de caso com pesquisa anterior realizada por Holmes (2015), utilizando os mesmos dados de dengue para o mesmo período, percebe-se que ao usar a binarização limita-se a informação diminuindo a eficiência do suporte a decisão. Sendo assim, esta nova abordagem permite usar toda a informação disponível conseguindo fornecer os níveis de prioridade, além de abrir um novo caminho na área da estatística espacial e oferece melhor concordância com o mapa de RIE.

O último estudo de caso trouxe dados da dengue mais recentes do ano 2017, esse estudo conseguiu avançar em relação aos anteriores pois utilizou votação por maioria *fuzzy* utilizando cardinalidade de conjuntos *fuzzy* e o *plurality voting fuzzy* utilizando cardinalidade de conjuntos *fuzzy*, além da construção do mapa por classes. Com a aplicação dos dados na arquitetura anterior proposta por Holmes (2015), foi possível afirmar que ao utilizar a teoria de conjuntos *fuzzy*, observou-se avanços e resultados que se aproximam do fenômeno estudado, conseguindo fornecer mapas de suporte a decisão mais próximo da realidade.

Considerando que decisão é um exercício constante da prática de saúde e todas atividades em saúde estão relacionadas com a busca e o uso da informação, quanto mais precisa for essa maior será a qualidade na tomada de decisão (MARIN, 2010). A nova arquitetura pode auxiliar a tomada de decisão de forma direta ou indireta dos gestores públicos de saúde, com intuito de melhorar o sistema de controle e prevenção de doenças, agregando a vantagem de ser menos restritivo utilizando a informação em sua totalidade.

Contudo, uma limitação encontrada para realização deste estudo consiste na incompletude dos dados notificados no Sinan, embora notificação compulsória e investigação dos casos suspeitos de dengue sejam obrigatórias. Esse problema é comum quando se utiliza

dados secundários de notificações e agravos, o que impede muitas vezes às pesquisas fornecerem resultados fidedignos, subestimando a situação real da população.

“A motivação é uma porta que se abre por dentro.”

Mario Sérgio Cortella

9 CONSIDERAÇÕES FINAIS

Neste trabalho foi proposta uma nova arquitetura utilizando regras de combinação *fuzzy* para métodos de aglomeração espacial aplicada à epidemiologia, sendo esta formalizada matematicamente, apresentando a flexibilidade de utilizar simultaneamente vários métodos de aglomeração espacial, como também a aplicação de qualquer regra de combinação. Vale ressaltar que esta abordagem é inovadora pois não foi encontrado nenhum estudo que aplicasse regras de combinação *fuzzy* para combinar de métodos de aglomeração espacial.

A aplicação desta nova arquitetura nos três estudos de casos, demonstraram sua funcionalidade e vantagens frente a arquitetura proposta anteriormente, já que a combinação dos métodos de aglomeração espacial adotando as regras de combinação *fuzzy*, ou seja, com a inclusão das informações em sua totalidade, gerou melhores resultados do que quando foram utilizados individualmente, tornando possível a obtenção de informações mais fidedignas e tomada de decisões mais precisas.

A utilização desta metodologia, abre um novo caminho na área de estatística espacial pois, sua utilização busca resultados mais precisos para uma melhor tomada de decisão. Tornando-se possível uma maior compreensão das populações vulneráveis ao acometimento e a localidade, facilitando assim a definição de políticas públicas para as áreas prioritárias e de transição, específicas para população daquela localidade. Pode-se também junto aos profissionais de saúde definir políticas de combate e prevenção à morbidade, tornando-se possível a diminuição do acometimento por Dengue.

O resultado da combinação das análises dos aglomerados espaciais, permitiu a visualização da distribuição espacial da Dengue em todos os municípios da Paraíba, demonstrando que os municípios localizados no agreste e sertão paraibano são mais vulneráveis ao acometimento da Dengue, direcionando os gestores para decisões que respeitem as particularidades de cada local.

Sendo assim, é possível afirmar que a combinação dos métodos de aglomeração espacial utilizando regras de combinação *fuzzy*, permite a inclusão das incertezas presentes no processo saúde-doença, preservando assim informações, possibilitando resultados mais fidedignos e tomada de decisões mais precisas.

Diante disso, a arquitetura foi construída de forma que possa ser utilizada de diferentes modos, utilizando as várias regras de combinação existentes e métodos de aglomeração diversos, podendo ser aplicada em outras regiões geográficas e em outros agravos, podendo

gerar estudos subsequentes. Uma limitação encontrada para realização deste estudo consiste na incompletude dos dados notificados, como discutido anteriormente.

Por fim, espera-se que estes resultados contribuam para o fortalecimento dos estudos em epidemiologia espacial, apontando a necessidade de sua utilização para uma maior compreensão das populações vulneráveis ao acometimento e a localidade que estão ocorrendo os agravos, norteando assim a definição de medidas para as áreas prioritárias com base nas necessidades específicas para cada localidade.

PRODUÇÃO ACADÊMICA

SÁ, L. R. DE; MELO, J. C. S.; NOGUEIRA, J. A.; MORAES, R. M. Combination of Spatial Clustering Methods Using Weighted Average Voting for Spatial Epidemiology. In: Guilherme A. Barreto; Ricardo Coelho. (Org.). *Fuzzy Information Processing*. 1ed. Switzerland: Springer, 2018, v. 831, p. 49-60.

SÁ, L. R.; MELO, J. C. S.; NOGUEIRA, J. A.; MORAES, R. M. Influencia do índice de Desenvolvimento humano (IDH) na distribuição espacial da Dengue. In: 18 Congresso Brasileiro dos conselhos de Enfermagem, 2015, João Pessoa. Anais do 18 Congresso Brasileiro dos conselhos de Enfermagem, 2015.

Artigos Submetidos:

SÁ, L. R.; SOARES, E. A. M. G.; FARIAS, A. D.; SANTIAGO, R.H. N.; BEDREGAL, B.; NOGUEIRA, J. A.; MORAES, R. M. A New Class of Rules for Combining Spatial Clustering Methods Based on Generalized Mixture Functions. *International Journal of Geographical Information Science*. Submetido em Abril/2019.

PINTO, M. M. P. S.; SÁ, L. R.; MEDEIROS, A. T.; NOGUEIRA, J. A.; MORAES, R. M.. Análise comparativa de métodos de aglomeração espacial para estudos epidemiológicos. *Revista Espaço e Geografia*. Submetido em Setembro/2018.

REFERENCIAS

ANSELIN, L. Spatial data analysis with GIS: an introduction to application in the social sciences. **National Center for Geographic Information end Anlisis**. University of California - Santa Barbara. August, 1992.

ARAÚJO, Kality Adja Medeiros de et al. Epidemiological study of scorpion stings in the Rio Grande do Norte State, Northeastern Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 59, 2017.

ASSUNÇÃO, R. M.; DICZMAL, A. R. a simulated annealing strategy for the detection of arbitrary shaped spatial clusters. **Computational Statistics and data analysis**, v. 45, p. 269-286, 2004.

BHATT, S. et al. The global distribution and burden of dengue. **Nature**. v. 496, p. 504–507 2013.

BASSANEZI, R. C., BARROS, L. C., OLIVEIRA, R. Z. G., LEITE, M. B. F. A disease evolution model with uncertain parameters In: 9th **IFSA Word Congress and 20th NAFIPS International Conference**, 2001, Vancouver. Proceedings - International Fuzzy Systems Association. IEEE - NAFIPS, 2001. v.3. p.1626 - 1630

BENMOKHTAR, V; B. HUET, Neural network combining classifier based on dempster-shafer theory for semantic indexing in video content, in: C. D. D. R. T. S. C. L. T. C. T J Cham, J Cai (Ed.), **Proceedings of the 13th International Conference on Multimedia Modeling**: 9-12 January 2007; Singapore, Berlin: Springer, 2007, pp. 196–205. doi:10.1007/978-3-54069429-8 20.

BESAG, J; NEWELL, J. The detection of clusters in rare diseases. **Journal of the Royal Statistic Society**, v. 154, p. 143-155, 1991.

BHOPAL, R. Concepts of epidemiology: an integrated introduction to the ideas, theories, principles and methods of epidemiology. **Oxford: Oxford University Press**; 2007.

BRASIL, Boletim Epidemiológico, Dengue, 2014.

BREVE, F. A.; PONTI, M. P.; MASCARENHAS, N. D. A. Combining methods to stabilize and increase performance of neural network-based classifiers, in: M. A. F. R. A C Frery (Ed.), **Proceding of the XVIII Brazilian Symposium on Computer Graphics and Image Processing: IEEE Computer Society**, p. 105–111, 2005.

CÂMARA, G.; DAVIS, C.; MONTEIRO, A.M., D'ALGE, J.C. Introdução á ciência da geoinformação. São José dos Campos: INPE, 2001.

CARVALHO, M. S.; SOUZA-SANTOS, R. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cad. saúde pública*, v. 21, n. 2, p. 361–378, 2005.

CASTRUCCI, B. Introdução à Lógica Matemática. Nobel, 1975.

COSTA, M.A.; ASSUNÇÃO, R.M. A fair comparison between the spatial scan and the Besag-Newell disease clustering tests. *Environmental and Ecological Statistics*, Springer, 1 v.12, p. 301-319, 2005.

COSTA, D. C. S. Tomada de decisão baseada em lógica Fuzzy e na distribuição espacial da mortalidade por acidentes de trânsito na cidade de João Pessoa – PB. 2011. Dissertação (Mestrado) – Pós-graduação em Modelos de Decisão e Saúde, Departamento de Estatística, Universidade Federal da Paraíba, João Pessoa, 2011.

DAI, L.; LIU, C. Research on remote sensing image of land cover classification based on multiple classifier combination, *Wuhan University Journal of Natural Sciences* 16 (4) (2011) 363–368. doi:10.1007/s11859011-0764-5.

DATTA, S.; PIHUR, V.; DATTA, S. An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data, *BMC Bioinformatics* 11 (2010) 1–11.

DOLFUD, O. O espaço geográfico. Rio de Janeiro: Bertrand Brasil, 1991.

DUIN, R.P.W.; TAX, D.M.J. Experiments with Classifier Combining Rules. **Pattern Recognition Group**, LNCS 1857, p.16-29. 2000.

ELLIOT, P.; WAKEFIELD, J. Spatial epidemiology: methods and applications. In: **D Briggs. Spatial epidemiology methods and applications**. Oxford: Oxford University Press;2000. p. 3-14.

FANTINATI, A. M. M. Demographic and epidemiological profile of dengue cases in central Goiania - Goiás: 2008 to March 2013. **Rev Tempus Actas Saúde Col**, 2013.

FARIAS, A. D. S. et al. A method of image reduction and noise reduction based on a generalization of ordered weighted averaging functions. arXiv:1601.03785, 2016.

GRACIA, E; LÓPEZ-QUÍLEZ, A.; MARCO, M.; LLADOSA, S.; LILA, M. A Epidemiologia Espacial da Violência do Parceiro Íntimo: As Vizinhanças Importam ?, **Revista Americana de Epidemiologia** , Volume 182, Edição 1, 1 de julho de 2015, Páginas 58–66, <https://doi.org/10.1093/aje/kwv016>.

GUZMAN, MARIA G.; GUBLER, DUANE J.; IZQUIERDO, ALIENYS; MARTINEZ, ERIC; HALSTEAD, SCOTT B. Dengue infection. **Nature Reviews Disease Primers**. 2016. 2. Macmillan Publishers Limited <https://doi.org/10.1038/nrdp.2016.5510.1038/nrdp.2016.55>

HOLMES, D. C. S. C. Uma nova arquitetura para combinação de aglomerados espaciais e sua aplicação em epidemiologia. 2015. Tese (Doutorado) – Pós-graduação em Modelos de Decisão e Saúde, Departamento de Estatística, Universidade Federal da Paraíba, João Pessoa, 2015.

HOLMES, D. C. S. C.; MORAES, R. M.; VIANNA, R. P. T. A Rule for Combination of Spatial Clustering Methods. *Patterns 2015: The Seventh international Conferences on Pervasive Patterns and Applications*.

HOLMES, D. C. S. C.; MORAES, R. M.; VIANNA, R. P. T. A new map for Combination of Spatial Clustering Methods. *Flins*, 2018.

IBGE. **Síntese de indicadores sociais 2010 uma análise das condições de vida da população brasileira**. Rio de Janeiro: IBGE, 2010.

JAIN, A. K et al. Statistical pattern recognition: A review, **IEEE Transaction on Pattern Analysis and Machine Intelligence**, v. 22, p. 4–37, 2000)

KEENAN, P. JANKOWSKI, P. Spatial Decision Support Systems: Three decades on. **Decision Support Systems**, 2018.

KITTLER, J.; HATEF, M.; DUIN, R. P. W.; MATAS, J. On Combining Classifiers. **IEEE Transaction on Pattern Analysis and Machine Intelligence**. v.20, n.3, p.226-239, 1998.

KOSKO, B. Counting with Fuzzy Sets .IEEE Transactions **On Pattern Analysis And Machine Intelligence**, v. PAMI-8, n. 4,1986.

KUNCHEVA, L.I. Combining Pattern Classifiers: Methods and Algorithms, **Wiley-Interscience**, New Jersey, 2004.

KULLDORFF, M. et al. An elliptic spatial scan statistic. **Statistics in Medicine**, v. 25, p. 3929-3943, 2006

KULLDORFF, M.; NAGARWALLA, N. Spatial disease cluster: detection and inference. **Statistics in Medicine**, v. 14, p. 799-810, 1995

LANTZ, B. Machine Learning with R. Birmingham: Packt Publishing, 2013

LATIF-SHABGAHI, G. R. A novel algorithm for weighted average voting used in fault tolerant computing systems. **Microprocessors and Microsystems**. 2004; 28(7):357–361.

LEE, G.; YAMADA, I.; ROGERSON, P. A. GeoSurveillance: Software for monitoring change in geographic patterns. State University os New York, Buffalo, 2007.

LI, C.S.; WANG, Y.; YANG, H. Combining Fuzzy c-Means Classifiers Using Fuzzy Majority Vote. **Fifth International Conference on Fuzzy Systems and Knowledge Discovery**, 2008.

LI, C.S.; WANG, Y.; YANG, H. Combining Fuzzy Partitions Using Fuzzy Majority Vote and KNN. **Journal of computers**, v. 5, n. 5, 2010.

LUCENA, S. E. DE F.; MORAES, R. M. Detection of space-time cluster of homicides by stabbing in the João Pessoa city, Paraíba state, Brazil. **Boletim de Ciências Geodésicas**, v. 18, n. 4, p. 605–623, 2012.

LUCENA, S.E.F.; MORAES, R.M. Análise do desempenho dos métodos SCAN e BESAG e NEWELL para identificação de conglomerados espaciais do dengue no município de João Pessoa entre os meses de janeiro de 2004 e dezembro de 2005. **Boletim de Ciências Geodésicas**, v. 15, n. 4, p. 544-561, 2009.

MARIN, H. F. Sistemas de informação em saúde: considerações gerais. **J. Health Inform.** Jan-v.2 , n. 1, p. 20-24, 2010

MEDRONHO, R. DE A.; BLOCH, K. V.; WERNECK, G. L. Epidemiologia. **Editora Atheneu**, 2009.

MELO, J.C.S.; MORAES, R.M.. Sistema Espacial de Suporte à Decisão para Gestão do Combate ao Dengue usando Lógica Fuzzy. **TEMA (São Carlos)**, São Carlos, v. 19, n. 3, p. 405-421, Dec. 2018. Available from
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2179-84512018000300405&lng=en&nrm=iso>. access
on 06 July 2019. <http://dx.doi.org/10.5540/tema.2018.019.03.0405>

MORA, M.; FORGIONNE, G.; GUPTA, J. Decision-making support systems: Achievements challenges, and trends. Hershey, **PA: Idea grup publishing**. 2002.

MORAES, R.M.; BANON, G.J.F.; SANDRI, S.A. Fuzzy expert systems architecture for image classification using mathematical morphology operators. **Information Sciences**, v. 142, n.1, p. 7-21, 2002.

MORAES, R. M.; MACHADO, L. S. A new class of assessment methodologies in medical training based on combining classifiers, in: Proceeding of the XII Safety, Health and Environment World Congress: 22-25 July 2012; São Paulo, COPEC, 2012, pp. 91–95.

MORAES, R. M.; NOGUEIRA, J. A.; SOUSA, A. C. A new architerture for a spatio-temporal decision support system for epidemilogical purposes. **Anais do Flins**, João Pessoa, 2014.

MOURA, F.R. Detecção de clusters espaciais via algoritmo SCAN multi-objeto. 2006. 51 f. dissertação (Mestrado) – Pós-graduação em Estatística, Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

NEIL, D. B.; MOORE, A. W.; SABHNANI, M; DANIEL, K. Detection of emerging space-time clusters. In: **Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery** (KDD '05), Chicago (USA), August 21-24, p. 218-227, 2005

NEIL, D. B.;MOORE, A. W.; SABHNANI, M. Morbidity and mortality weekly report. In:____.v. 545, p.197, cap. Detecting Elongated Disease Clusters 2005.

OPENSHAW, S.; CRAFT, A.W.; CHARLTON, M.; BIRCH, J.M. Investigation of leukaemia clusters by use of a geographical machine. *Lancet*, p. 272-273, 1988.

PASSINO, K. M. Fuzzy control. Menlo Park, Calif: Addison-Wesley, 1998

PONTI JUNIOR, M. P. Combinação de múltiplos classificadores para identificação de materiais em imagens ruidosas. 2004. Dissertação (Mestrado), Universidade Federal de São Carlos, São Carlos, 2004.

RESENDES, A. P. C. Sensoriamento remoto e modelagem espacial nos estudos dos processos endêmicos-epidêmicos em áreas urbanas: o caso da dengue no município de Niterói - Estado do Rio de Janeiro. 2010. Tese (Doutorado) – Pós-graduação em Ciência na área de Saúde Pública, Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2010.

ROGERSON, P. A. The detection of clusters using a spatial version of the chi-square goodness-of-fit statistic. **Geographical Analysis**, v. 31, 1999.

ROGERSON, P. A. Spatial monitoring of geographic patterns: na aplicação to crime analysis. *Computes, Environment and Urban Systems*, v. 25, p. 539-556, 2001.

ROTHMAN, K.; LASH, T.; GREENLAND, S. *Modern Epidemiology*. Wolters Kluwer (2012).

ROUQUAYROL, M.Z. *Epidemiologia e saúde*. São Paulo, Medsi, 1986, p. 1.

SÁ, L.; NOGUEIRA, J. DE; MORAES, R. Modelo de decisão sobre o perfil demográfico para o controle da tuberculose usando lógica Fuzzy. **Revista Eletrônica de Enfermagem**, v. 17, n. 2, p. 223-37, 30 jun. 2015.

SÁ, L. R. DE; MELO, J. C. S. ; NOGUEIRA, J. A. ; MORAES, R. M. . Combination of Spatial Clustering Methods Using Weighted Average Voting for Spatial Epidemiology. In: Guilherme A. Barreto; Ricardo Coelho. (Org.). *Fuzzy Information Processing*. 1ed. Switzerland: Springer, 2018, v. 831, p. 49-60.

SÁ, L. R.; MELO, J. C. S.; MORAES, R. M. ; NOGUEIRA, J. A. . Influência do índice de Desenvolvimento humano (IDH) na distribuição espacial da Dengue. In: 18 Congresso Brasileiro dos conselhos de Enfermagem, 2015, João Pessoa. **Anais do 18 Congresso Brasileiro dos conselhos de Enfermagem**, 2015.

SANDERSON, C.; GRUEN, R. *Analytical Models for Decision Making*. **London: Open University Press**, 2006.

SAN PEDRO, Alexandre et al . Condições particulares de produção e reprodução da dengue em nível local: estudo de Itaipu, Região Oceânica de Niterói, Rio de Janeiro, Brasil. *Cad. Saúde Pública*, Rio de Janeiro , v. 25, n. 9, p. 1937-1946, set. 2009 . Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2009000900008&lng=pt&nrm=iso>. acessos em 06 jul. 2019. <http://dx.doi.org/10.1590/S0102-311X2009000900008>.

SANTOS, Claudia Benedita dos et al . Utilização de um Sistema de Informação Geográfica para descrição dos casos de tuberculose. **Bol. Pneumol. Sanit.**, Rio de Janeiro , v. 12, n. 1, p. 07-12, abr. 2004 .

SKALINSKI, L.M.; COSTA, M.C.; TEIXEIRA, M.G.L. Contribuições da análise espacial para a compreensão da dinâmica de transmissão da dengue: revisão integrativa. **J. Health Biol Sci**. v. 7, n. 1, p. 53-63, 2019.

SOARES, D. C. S. C. **MODELO DECISÓRIO ESPACIAL PARA A REDUÇÃO DA MORTALIDADE INFANTIL: uma discussão no contexto da ruralidade na Paraíba**. 2018. Tese (Doutorado) – Pós-graduação em Modelos de Decisão e Saúde, Departamento de Estatística, Universidade Federal da Paraíba, João Pessoa, 2018.

SOUZA, W. V. et al. Tuberculosis en asentamientos intraurbanos: un enfoque Bayesiano. **Tropical Medicine & International Health**, v. 12, n. 3, p. 323–330, 2007.

SUEN, C. Y.; NADAL, C.T. A.; MAI, R.L; LAM, L. Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts. *Frowiers in Handwriting*

Recognition, C. Y. Suen, Ed., in Proc. Int. Workshop on Frontiers in Handwriting Recognition, Montreal, Canada, Apr, v.2, n.3, p.131-143, 1990.

SUGUMARAN, R.; DEGROOTE, J. Spatial Decision Support Systems. CRC Press. 2011.

SUSSER, M.; SUSSER, E. Choosing a future for epidemiology: I. Eras and paradigms. *Am J Public Health*. May; v. 86, n. 5, p. 668-73, 1996.

TANGO, T. A class of tests for detecting ‘general’ and ‘focused’ clustering of rare diseases. *Statistics in Medicine*, v. 14, p. 1307-1315, 1995.

TANGO, T.; TAKAHASHI, K. A flexibly shaped spatial scan statistic for detecting clusters. **International Journal of Health Geographics**, v. 4, n. 11, 2005.

WEBB, A. R. Statistical Pattern Recognition. ed.2. QinetiQ Ltd., Malvern, UK. 2002.

WHO. Dengue Fever. Epidemiologic. 2015.

WHO. Dengue Fever. Epidemiologic. 2016.

WORBOYS, M. F. GIS: A computing Perspective. London: Taylor and Francis, 1995.

XU, L.; KRZYZH A.; SUEN, C.Y. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. **IEEE Transaction on Systems, Man, and Cybernetics**, v.22, n.3, p.418-435, 1992.

ZADEH, L. A. Fuzzys Sets. 1965

ZADEH, L. A. A computational approach to fuzzy quantifiers in natural languages. **Comput. Math. with Applications**, v. 9, n. 1, p. 149-184, 1983.

ZADEH, K. S. Fuzzy health, illness, and disease. **The Journal Of Medicine And Philosophy**, v. 25, n. 5, p. 605–638, out. 2000.

ZHOU, Z. H. Ensemble Methods Foundations and Algorithms. **Chapman & Hall/CRC Machine Learning & Pattern Recognition Series**, 2012.