

Otaciana Gleisa Rezende Santiago

Investigação para melhorias na abordagem de comitês homogêneos e heterogêneos sobre uma base biométrica de assinatura através de análises experimentais.

Natal-RN

2012

Otaciana Gleisa Rezende Santiago

Investigação para melhorias na abordagem de comitês homogêneos e heterogêneos sobre uma base biométrica de assinatura através de análises experimentais.

Monografia apresentada na disciplina Trabalho de Diplomação, como requisito parcial para a conclusão do curso de Bacharelado em Ciência da Computação na Universidade do Estado do Rio Grande do Norte

Orientador: Prof. Ds. Isaac de Lima Oliveira Filho

UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE - UERN
CAMPUS AVANÇADO DE NATAL

Natal-RN

2012

Resumo

O reconhecimento biométrico é um método de autenticação de usuários que se diferencia dos métodos convencionais pela forma de como é realizada a identificação. Eles utilizam características corporais únicas que podem ser características físicas ou comportamentais. Tipicamente o processo para realizar a autenticação de uma pessoa é baseado em duas fases. Inicialmente é realizada a coleta e armazenamento da informação biométrica e depois é realizada uma verificação na informação fornecida e comparada a base de dados que fornece a identificação do indivíduo com unicidade. As informações são transformadas em dados numéricos e então são utilizadas técnicas de Inteligência artificial. Existem vários algoritmos de Aprendizado de máquina que podem ser utilizados para o reconhecimento biométrico, mas pode ser destacada a técnica conhecida como multiclassificadores ou comitês que é a utilização de vários algoritmos combinados. Constata-se na literatura que a utilização desta técnica de combinação de algoritmos sobre bases de dados, garante um percentual maior de classificação. Este trabalho se propõe a utilizar a combinação dos algoritmos IBK, MLP e J48, buscando encontrar melhores taxas de aprendizado em uma base biométrica de assinaturas e visa encontrar uma configuração para o comitê que gere as melhores taxas de aprendizado. É utilizado o método de validação cruzada para assegurar que uma medida verdadeira esteja sendo estimada pelos classificadores. Depois esses comitês são submetidos aos metaclassificadores Naive bayes, SMO, IBK, soma e voto.

Palavras-Chave: Biometria; Aprendizado de Máquina; Comitê.

Abstract

The biometric recognition is a method of authentication of users which is different of the common methods for the way how it performs the recognition. It uses unique bodily characteristics which can be physical or behavior characteristics. Usually the process to perform the authentication of a person is based in two phases. On a first one is made a collection of biometric information and after the collected information is verified and compared to a data base containing the unique individual information. The collected information is changed to numeric data for being used in artificial intelligence techniques. There are many algorithms used to make biometric recognition but among them can be mentioned the use of some algorithms combined. This technique is known as multi-classifiers or ensemble. This work proposes to make a combination of IBK, MLP and J48 algorithms as a way to improve the learning on a biometrical base of signatures. It aims to find a committee setting that generates best accuracy rates. The cross validation method is used to ensure that the right measure is being estimated by the classifiers. After the committees are subjected to the meta-classifiers Naive Bayes, SMO, IBK, Soma and Voto.

Keywords: biometrics; machine learning; ensemble.

Agradecimentos

Eu não poderia começar os agradecimentos de outra forma, senão agradecendo a Deus que me conduz e me leva nos braços sempre que não tenho forças para caminhar e a Maria, mãe de Deus e minha, que escutou minhas orações e intercede por mim, incessantemente.

À minha família, especialmente a minha mãe Geruzia, por está sempre ao meu lado me dando apoio e por suas orações. Ao meu pai Otacílio e ao meu irmão Oderley, pelos conselhos e torcida.

Ao meu namorado Anchieta pelo amor, amizade, compreensão, incentivo e dedicação.

Ao meu orientador, Isaac de Lima Oliveira Filho, pela confiança e paciência com que me acompanhou no desenvolvimento desse trabalho.

Enfim, agradeço a todos que me ajudaram na realização desse sonho, torcendo por mim.

Sumário

Lista de Figuras	p. vii
Lista de Tabelas	p. ix
Lista de Abreviaturas e Siglas	p. x
1 Introdução	p. 1
1.1 Objetivos	p. 3
1.2 Estrutura do documento	p. 3
2 Bases Biométricas	p. 4
2.1 Funcionamento da verificação biométrica	p. 4
2.2 Tipos de sistemas biométricos informatizados	p. 5
2.2.1 Sistemas Biométricos Fisiológicos	p. 5
2.2.1.1 Impressão Digital	p. 6
2.2.1.2 Reconhecimento Facial	p. 6
2.2.1.3 Reconhecimento da Íris	p. 7
2.2.2 Sistemas Biométricos Comportamentais	p. 8
2.2.2.1 Dinâmica de Voz	p. 8
2.2.2.2 Verificação de Assinatura	p. 8
2.2.3 Base de Dados	p. 10
3 Aprendizado de Máquina	p. 11
3.1 Tipos de Aprendizado	p. 12

3.1.1	Aprendizado Supervisionado	p. 12
3.1.2	Aprendizado Não-Supervisionado	p. 13
3.2	Algoritmos Individuais	p. 13
3.2.1	Algoritmo Árvore de decisão - J48	p. 13
3.2.2	Algoritmo Multi-Layer Perceptron - MLP	p. 15
3.2.3	Algoritmo K-Vizinho Mais Próximos - IBK	p. 17
3.3	Sistema Multiclassificador	p. 18
3.3.1	Combinando Classificadores Homogêneos	p. 20
3.3.1.1	Bootstrap Aggregating (Bagging)	p. 20
3.3.1.2	Boosting	p. 21
3.3.2	Combinando Classificadores Heterogêneos	p. 22
3.3.2.1	Stacking	p. 22
3.3.2.2	StackingC	p. 23
4	Metodologia dos Experimentos	p. 24
4.1	Conjunto de dados	p. 24
4.1.1	Atributos das assinaturas	p. 25
4.1.1.1	Tempo de execução	p. 26
4.1.1.2	Suspensões da caneta	p. 26
4.1.1.3	Largura da imagem	p. 26
4.1.1.4	Altura da imagem	p. 27
4.1.1.5	Razão entre altura e largura	p. 27
4.1.1.6	Velocidade horizontal média	p. 27
4.1.1.7	Velocidade vertical média	p. 28
4.1.1.8	Cruzamentos verticais da caneta no ponto central	p. 28
4.1.1.9	Momentos de Invariância	p. 28
4.1.2	Metodologia da Validação	p. 30

4.1.2.1	Validação Cruzada	p. 30
4.1.3	Experimentos	p. 31
5	Resultados dos Experimentos	p. 32
5.1	Resultados dos Multi-Classificadores Homogêneos	p. 32
5.2	Resultados dos Multiclassificadores Heterogêneos	p. 34
5.3	Melhores resultados dos MultiClassificadores	p. 35
5.3.1	Resultados do Metaclassificador SMO	p. 35
5.3.2	Resultados do Metaclassificador Ibk	p. 37
5.3.3	Resultados do Metaclassificador NB	p. 37
5.3.4	Resultados do Metaclassificador Soma	p. 38
5.3.5	Resultados do Metaclassificador Voto	p. 39
5.3.6	Resultados dos Comitês rodando individualmente	p. 41
5.3.7	Discussão dos resultados	p. 42
6	Conclusões e trabalhos futuros	p. 46
6.1	Trabalhos Futuros	p. 46
	Apêndice A – Comitês que apresentaram os melhores resultados	p. 48
	Referências	p. 64

Lista de Figuras

1	Funcionamento dos Sistemas Biométricos	p. 5
2	Impressão Digital	p. 6
3	Reconhecimento Facial	p. 7
4	Reconhecimento da Íris	p. 7
5	Espectrograma de timbre de voz	p. 9
6	Mesa Digitalizadora	p. 9
7	Diagrama de blocos da aprendizagem Supervisionada	p. 12
8	Diagrama em blocos da aprendizagem Não-Supervisionada	p. 13
9	Árvore de decisão para duração de uma viagem	p. 14
10	Modelo do neurônio proposto por McCulloch e Pitts	p. 16
11	Exemplo de uma RNA multicamadas	p. 16
12	Exemplo de um hiperplano ótimo para padrões linearmente separáveis .	p. 17
13	Exemplo de classificação do método KNN	p. 18
14	Processo de combinação de decisão de um comitê classificador	p. 19
15	Distribuição da amostra por gêneros	p. 25
16	Distribuição da população estudada por faixa etária	p. 25
17	Formulário de coleta	p. 26
18	Assinatura original e projeção do ponto vertical central na assinatura .	p. 28
19	Média dos comitês homogêneos para configuração inicial dos valores dos parâmetros.	p. 43
20	Média dos comitês homogêneos para configuração com valores dos parâmetros menores.	p. 43

21	Média dos comitês homogêneos para configuração com valores dos parâmetros maiores.	p. 44
22	Média dos comitês heterogêneos para configuração inicial dos valores dos atributos.	p. 44
23	Média dos comitês heterogêneos para configuração com valores dos atributos menores.	p. 45
24	Média dos comitês heterogêneos para configuração com valores dos atributos maiores.	p. 45

Lista de Tabelas

1	Definições dos Momentos	p. 30
2	Média dos comitês homogêneos para o experimento 1	p. 33
3	Média dos comitês homogêneos para o experimento 2	p. 33
4	Média dos comitês homogêneos para o experimento 3	p. 33
5	Média dos comitês heterogêneos para o experimento 1	p. 34
6	Média dos comitês heterogêneos para o experimento 2	p. 34
7	Média dos comitês heterogêneos para o experimento 3	p. 34
	p. 36	
9	Resultados do Metaclassificador Ibk	p. 38
10	Resultados do Metaclassificador Naive Bayes	p. 39
11	Resultados do Metaclassificador Soma	p. 40
12	Resultados do Metaclassificador Voto	p. 41
13	Resultados individual	p. 42
14	Configurações dos comitês com três algoritmos de AM que obtiveram os melhores resultados para Taxa de Aprendizado e Desvio Padrão	p. 48
15	Configurações dos comitês com seis algoritmos de AM que obtiveram os melhores resultados para Taxa de Aprendizado e Desvio Padrão	p. 51
16	Configurações dos comitês com doze algoritmos de AM que obtiveram os melhores resultados para Taxa de Aprendizado e Desvio Padrão	p. 55

Lista de Abreviaturas e Siglas

K-NN - *K-Nearest Neighbor*

AD - *Árvores de Decisão*

RNA - *Redes Neurais Artificiais*

WEKA - (

1 *Introdução*

Atualmente um dos maiores problemas enfrentados em termos de segurança na informática é a autenticação de usuários. Na maioria dos sistemas de informação, essa autenticação é garantida através de senhas alfanuméricas, ou até mesmo por cartões com chips ou com dispositivos magnéticos, porém essas formas de autenticação são bastante vulneráveis em termo de segurança, pois permitem acesso de terceiros a esses sistemas. Devido a estes problemas, técnicas de autenticação baseadas em características biométricas vêm sendo utilizadas para garantir a autenticidade dos usuários com maior confiabilidade (HEINEN; S., 2002).

A biometria é a ciência que realiza a medição e a análise de dados biológicos, e utiliza essas características como mecanismos de identificação. O estudo deste campo teve início no século XX, e é dividido em duas classes principais: Fisiológicas (que inclui impressão digital, reconhecimento facial, geometria da mão e de reconhecimento da íris) e Comportamentais (verificação de assinatura, dinâmica de digitação e voz)(ROMANO, 2010).

A técnica de autenticação baseada em características biométricas físicas apesar de garantir a confiabilidade a respeito da autenticidade do usuário, apresenta algumas desvantagens, como o custo elevado do equipamento de hardware necessário para a autenticação e o alto grau de intrusão, que gera desconfortos aos usuários (HEINEN; S., 2002).

Dos métodos biométricos o que apresenta maiores vantagens é o método de verificação de assinatura, pois além de garantir a confiabilidade tem a vantagem de os usuários já estarem acostumados a utilizar assinaturas como uma forma de autenticação, e assim sentem-se mais seguros e confortáveis em relação ao seu uso. O grau de intrusão apresentado pelos sistemas de autenticação de assinaturas é baixo, e o hardware necessário tem um custo bastante acessível (HEINEN; S., 2002).

Mas apesar das vantagens apresentadas, na prática essa técnica apresenta um grande problema em relação ao ponto de vista computacional, devido à grande variabilidade que

ocorre entre assinaturas de uma mesma pessoa. Como solução métodos baseados em Inteligência Artificial (IA) e Aprendizagem de Máquina (AM) são utilizados.

Em métodos de AM o sistema computacional é programado para aprender ou induzir a partir de experiências passadas. Neste trabalho será utilizado três tipos de métodos de aprendizagem: K-NN (*K-Nearest Neighbor*), AD (*Árvores de Decisão*), RNA (Redes Neurais Artificiais). Todos são baseados no método de aprendizado supervisionado¹ diferenciando um do outro pela maneira que é realizado o treinamento. Existem também técnicas de aprendizado de máquina que são consideradas não-supervisionadas², é o caso dos algoritmos de clusterização como o K-means (FACELI, 2011).

No entanto, existem outras técnicas de otimização da utilização destes classificadores, que neste caso podem ser agrupados. Esta técnica é conhecida como comitê *ensemble*, que podem ser constituídos de classificadores do mesmo tipo (comitês homogêneos), alterando-se apenas suas variáveis estruturais, como também conter em sua estrutura classificadores de tipos diferentes (comitês heterogêneos). Constata-se na literatura que a utilização destes comitês sobre bases de dados, garante um percentual maior de classificação (NASCIMENTO, 2009).

Para realizar o treinamento de uma rede frequentemente é utilizada uma ferramenta de mineração de dados. Uma das ferramentas mais utilizadas é a WEKA (WAIKATO, 1993)(Waikato Environment for Knowledge Analysis) por ser um software livre, ter uma interface bastante simples e agregar um grande conjunto de algoritmos de classificação e ferramentas de pré-processamentos dos dados. Esta ferramenta também é bastante utilizada no meio acadêmico, favorecendo o compartilhamento de informações aos seus usuários. Além disso, desenvolvida em JAVA por pesquisadores da Universidade de Waikato na Nova Zelândia pode acessar dados oriundos de bancos de dados ou através da chamada de arquivos de dados próprios (PIMENTA *et al.*, 2009). Devido a essas vantagens, a ferramenta foi escolhida entre outras ferramentas de mineração existentes, como por exemplo, a Cubist (CUBIST, 1998).

¹Tipo de aprendizado de máquina baseado em experiências acumuladas contidas no conjunto de treinamento, ou seja, há uma classificação nos padrões

²Não existe a presença de uma saída desejada, nem a presença de um professor, ou seja, os padrões não apresentam sua classificação no conjunto de dados

1.1 Objetivos

O objetivo desse trabalho é analisar e comparar resultados dos comitês homogêneos e heterogêneos sobre uma base biométricas de assinatura utilizando os classificadores: IBK (*K-NN*), J48 (*AD*) e *Multi-Layer Perceptron* (MLP). Este trabalho ainda pode ser subdividido nos seguintes objetivos específicos:

- Apresentar os algoritmos de aprendizado de máquina, como também definir e encontrar os melhores comitês para a base biométrica.
- Comparar a acurácia e demais índices, como erros e desvio padrão dos algoritmos quanto ao desempenho sobre as bases.
- Utilizar a ferramenta WEKA para executar os classificadores, testar e validar os resultados.

1.2 Estrutura do documento

Este trabalho se encontra organizado em 5 capítulos que estão sub-divididos em seções. Cada capítulo procura contribuir de alguma forma para o esclarecimento da pesquisa.

O capítulo 2, aborda os conceitos sobre Bases biométricas, tipos de biometria e o funcionamento desses métodos biométricos.

O capítulo 3, concentra-se nos conceitos de Aprendizado de Máquina, tipos de aprendizado, Classificadores e Comitês, assim como o funcionamento dos classificadores individuais e combinações de classificadores.

O capítulo 4, apresenta a metodologia utilizada nos experimentos, bem como a base de assinaturas.

No capítulo 5, são expostos os testes e análise dos resultados obtidos de cada classificador e Comitê através da utilização da base de assinaturas na ferramenta WEKA.

No capítulo 6, são apresentadas as conclusões obtidas a partir dos experimentos que foram realizados e as perspectivas futuras de outros trabalhos que possivelmente poderão utilizar este como consulta.

Por fim, são apresentadas a bibliografia utilizada neste trabalho e o apêndice, onde são apresentadas as configurações utilizadas pelos algoritmos com os valores de cada parâmetro.

2 Bases Biométricas

O termo biometria do grego bios (vida) + metron (medida), é a ciência que realiza a medição e a análise de dados biológicos em mecanismos de identificação e autenticação. A biometria se baseia no fato do ser humano possuir características corporais únicas e que são relativamente estáveis, por exemplo, impressões digitais, íris e retinas, padrões de voz, assinaturas e etc (FURTADO, 2002).

Segundo Martins (2007), essas características são utilizadas na autenticação biométrica porque apresentam as seguintes propriedades: universalidade (todas as pessoas têm a característica); singularidade (duas pessoas não têm a mesma característica); permanência (a característica é invariante com o tempo); contabilidade (a característica pode ser medida quantitativamente).

Cada biometria tem os seus pontos fortes e limitações, e cada tipo de biometria é adequada para uma determinada aplicação. Os diferentes tipos de biometria estão inseridos em duas classes principais de acordo com suas características biométricas e apresentam o mesmo funcionamento básico para realização da verificação biométrica, que será abordado na próxima seção.

2.1 Funcionamento da verificação biométrica

Os sistemas biométricos são considerados sistemas de reconhecimento de padrões com um propósito específico. O reconhecimento de padrões é a capacidade de distinguir padrões e separá-los em diferentes classes ou categorias (ROMANO, 2010).

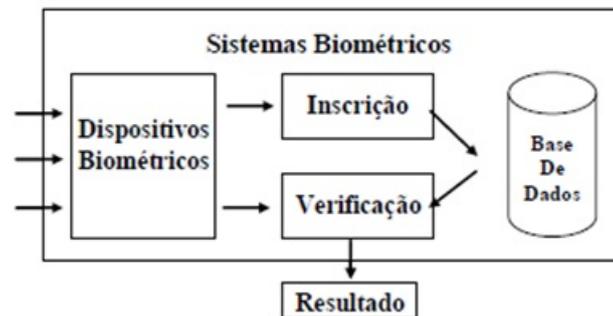
Esses padrões são utilizados com o objetivo de autenticar pessoas de forma segura. O procedimento para realizar essa verificação biométrica e verificar a autenticidade de uma pessoa segue os seguintes passos (FURTADO, 2002):

- Na primeira fase é realizado o registro das informações do usuário no sistema biométrico, esse registro é realizado por sensores que digitalizam e capturam as infor-

mações biométricas e depois armazena na base de dados.

- Na segunda fase, é realizada uma verificação na informação fornecida e comparada a base de dados que fornece a identificação do indivíduo com unicidade. Esse procedimento é ilustrado na Figura 1.

Figura 1: Funcionamento dos Sistemas Biométricos



Fonte: (ROMANO, 2002)

De acordo com (SILVA; M., 2007), esses sistemas utilizam basicamente três componentes para realizar o processo de identificação ou autenticação:

- Um sensor, que detecta a característica que está sendo usada para a identificação.
- Um computador, com um aplicativo que recebe e armazena as informações.
- Um software, que analisa as características e as traduz para um gráfico ou código, fazendo as comparações.

2.2 Tipos de sistemas biométricos informatizados

Alguns dos principais sistemas informatizados de verificação biométrica serão apresentados nessa seção. Esses sistemas são classificados como sistemas Fisiológicos ou sistemas Comportamentais, dependendo da característica biométrica, conforme dito na seção anterior.

2.2.1 Sistemas Biométricos Fisiológicos

São sistemas que utilizam as técnicas relacionadas às formas do corpo. Como por exemplo, a impressão digital, reconhecimento facial e reconhecimento da Iris.

2.2.1.1 Impressão Digital

A biometria de digital é amplamente conhecida como um método preciso de identificação e verificação biométrica, pois as impressões digitais são únicas para cada pessoa e para cada dedo. De acordo a CBA (Consultores Biométricos Associados) a maior parte dos sistemas de digitais analisa pequenos atributos únicos na imagem da digital, como os contornos das linhas papilares ou bifurcações. Outros sistemas de impressões digitais analisam os pequenos poros no dedo, a densidade da imagem digital ou a distância entre as linhas papilares. Na Figura 2 pode ser visto um exemplo de uma impressão digital.

Figura 2: Impressão Digital



Fonte: SILVA; M., 2007

Essa técnica de biometria é sem dúvida a mais utilizada atualmente, por apresentar um nível de aceitação bastante satisfatório entre os usuários e pelo fato da impressão digital ser usada há muito tempo como forma de autenticação.

Porém, segundo Magalhães (2003), esta tecnologia é, entre as biometrias físicas, a de menor confiabilidade, pois fatores como sujeira, corte, idade, sexo, etnia e a força ou posição que o dedo é colocado no scanner podem distorcer a imagem e não ser reconhecida.

2.2.1.2 Reconhecimento Facial

Segundo ROMANO (2002), detecção de rostos é a determinação da existência ou não de um rosto na imagem e uma vez encontrado este objeto, sua localização deve ser apontada através de um enquadramento ou retornando as suas coordenadas dentro da imagem. Conforme mostrado na Figura 3.

O processo tem início com a captura de uma imagem, seguida da detecção de um rosto que será comparada com modelos armazenados numa base de dados, complementada com a análise da cor da pele, detecção de linhas ou ainda de um modelo híbrido (THIAN, 2001).

Os principais problemas encontrados nessa técnica é a orientação da cabeça e condições que podem alterar a aparência como uso de óculos, maquiagem, barba e envelhecimento.

Figura 3: Reconhecimento Facial



Fonte: SILVA; M., 2007

Essa técnica vem sendo aplicada em sistemas de vigilância, controle de acesso, definições automáticas de perfis, expressões faciais, como raiva, tristeza ou alegrias nos usuários (SILVA; M., 2007).

2.2.1.3 Reconhecimento da Íris

O reconhecimento do indivíduo através da íris é uma tecnologia relativamente nova e tem se mostrado estável e precisa. Esta tecnologia envolve a análise do anel colorido que cerca a pupila do olho humano e é a menos intrusiva de todas, funcionando mesmo com óculos postos (LIU S. E SILVERMAN, 2001). Na Figura 4, são mostrados alguns exemplos de íris.

Para adquirir uma imagem da íris utilizamos câmeras monocromáticas. Para extrair as características, é localizada dentro da imagem monocromática a pupila (que é centralizada), o padrão da íris é isolado da pupila e demodulado para extração de sua informação de fase. O processo de comparação calcula uma medida da similaridade por meio da distância de Hamming normalizada, um método que simplesmente calcula a quantidade da divergência de bits entre as codificações (ROMANO, 2002).

Figura 4: Reconhecimento da Íris



Fonte: ROMANO, 2010

Essa técnica é a que apresenta as menores taxas de erro de reconhecimento, devido o fato da Iris nunca mudar durante toda vida do individuo, porém apresenta custos elevados e problemas relacionados à distância em que o olho está em relação ao equipamento usado

para fazer a biometria e também a doenças que o olho possa apresentar como catarata, conjuntivite ou alergia o reconhecimento será prejudicado (VALADAO, 2008).

2.2.2 Sistemas Biométricos Comportamentais

Os Sistemas Biométricos Comportamentais são sistemas que analisam as técnicas relacionadas ao comportamento de uma pessoa, a autenticação de um indivíduo é realizada através de observações de como ele executa determinadas ações. São exemplos desse tipo de biometria os sistemas de verificação de assinatura, dinâmica de digitação e voz.

2.2.2.1 Dinâmica de Voz

A voz é uma característica única em razão do formato das cavidades vocais e da forma que se movimenta a boca ao falar. Segundo Oliveira (2007), a biometria de voz consiste em um processo de identificação através da avaliação de uma elocução do usuário, comparando-a a amostras (também chamadas *voiceprints*) previamente guardadas em uma base de dados.

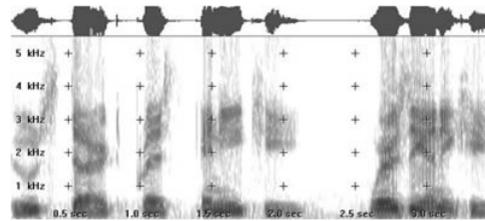
O programa de identificação faz uma comparação entre produções de uma mesma fala e uma análise dos padrões harmônicos. Essa característica é importante para evitar fraudes, como o uso de gravações.

Esse tipo de biometria é útil para aplicações em que as pessoas precisem ter acesso à informações ou que precisem passar informações sem estar fisicamente presentes baseadas e em aplicações de telefonia. A mesma apresenta vantagens como baixo custo do equipamento de hardware, uma vez que não necessita de dispositivos especiais, podendo fazer uso apenas de um telefone ou de um microfone conectado a um computador. Entretanto, barulhos e variações comportamentais humanas que corresponde a variações na maneira de falar devido a condições emocionais extremas ou mesmo a um simples resfriado podem afetar o desempenho dos sistemas de identificação de fala. Na figura 5, pode-se observar um exemplo de espectograma para o timbre da voz.

2.2.2.2 Verificação de Assinatura

A assinatura pode ser entendida como uma sequência lógica de movimentos previamente gravados no cérebro que, praticamente, são impossíveis de serem reproduzidos. Em um sistema de verificação de assinaturas a autenticação pode ser realizada de duas formas diferentes: autenticação off-line, onde a assinatura é feita pelo usuário em uma

Figura 5: Espectrograma de timbre de voz



Fonte: SILVA; M., 2007

folha de papel, e posteriormente digitalizada em um scanner e enviada para o sistema que faz a autenticação e autenticação on-line, onde a assinatura é feita diretamente sobre um dispositivo especial de hardware, como uma mesa digitalizadora ou um tablet. A Figura 6, apresenta um exemplo de mesa digitalizadora .

Figura 6: Mesa Digitalizadora



Fonte: SILVA; M., 2007

O reconhecimento on-line, apresenta diversas vantagens em relação ao off-line, como maior riqueza de informações. Pois, além das características visuais da assinatura, é possível obter informações temporais e dinâmicas como pressão da caneta, velocidade, identificação dos movimentos da caneta, perda de contato com o papel, entre outras (CAVALCANTE *et al.*, 2005).

Dos métodos biométricos, o método de verificação de assinatura é o que apresenta maiores vantagens, pois além de garantir a segurança tem a vantagem de os usuários já estarem acostumados a utilizar assinaturas como uma forma de autenticação, e assim sentem-se mais seguros e confortáveis em relação ao seu uso. O grau de intrusão apresentado é baixo, e o hardware necessário tem um custo bastante acessível. Mas por ser um método que utiliza principalmente características comportamentais este método apresenta um grande problema em relação a variabilidade que ocorre entre assinaturas de uma mesma pessoa e o estado emocional, que pode interferir na assinatura (HEINEN; S., 2002).

O processo de autenticação de assinaturas, assim como os outros métodos de reconhe-

cimento biométrico, são tipos específicos de reconhecimento de padrões, onde a resposta do sistema deve indicar apenas se o padrão apresentado pertence ou não a classe (no caso do sistema biométrico de assinaturas, o sistema deve indicar assinatura verdadeira/falsa de um determinado usuário).

2.2.3 Base de Dados

Nesse trabalho, será analisada uma base biométrica de assinaturas, no qual será apresentada detalhadamente no capítulo 4. Essa base foi coletada pela Universidade de Kent em Canterbury no Reino Unido, onde foram coletadas assinaturas de 359 pessoas e analisadas características como idade, sexo e destreza manual, de cada pessoa.

3 *Aprendizado de Máquina*

No segundo capítulo, foram apresentados vários métodos de realizar a identificação de um indivíduo através da análise de atributos biométricos. Esses métodos surgiram como solução para problemas encontrados na autenticação de usuários por meios convencionais como senhas alfanuméricas. Porém, estes atributos precisam ser transformados em dados numéricos, binários ou coordenadas e submetidos a métodos de AM e IA.

AM (Aprendizado de Máquina) é uma área de pesquisa que utiliza técnicas para realizar a classificação de padrões de forma automática. Essas técnicas encontram a partir de experiências passadas, uma hipótese capaz de resolver o problema que se deseja tratar (FACELI, 2011).

O termo AM compreende diversas definições e conceitos, que na maioria das vezes convergem para um mesmo significado. Uma dessas definições é apresentada por Mitchell (MITCHELL, 1997), que define AM como a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência. Outra definição é proposta por Craven (CRAVEN; SHAVLIK, 1994) onde diz que, AM é o ramo da IA que desenvolve métodos capazes de aprender um conceito.

O AM utiliza métodos, técnicas ou sistemas computacionais para realizar um aprendizado sobre um determinado conceito ou tarefa. Existem diversas aplicações que utilizam o AM para realizar um aprendizado como solução de problemas, entre as quais podemos citar o gasto de energia elétrica de uma determinada cidade, reconhecimento de palavras e faces, detecção de uso fraudulento de cartões de crédito e identificação de usuários em sistemas de informática.

Além do grande número de aplicações que utilizam o AM, outros fatores contribuem para o crescimento dessa área, como o desenvolvimento de algoritmos cada vez mais robustos e eficientes (FACELI, 2011).

Baseados neste contexto, nesta seção serão apresentados os principais tipos de aprendizado de Máquina, bem como exemplos de algoritmos e combinações de algoritmos (co-

mitês) que serão utilizados nesse trabalho.

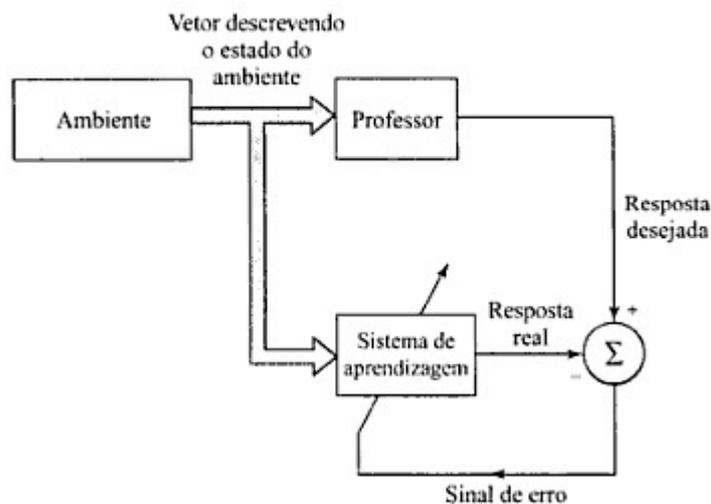
3.1 Tipos de Aprendizado

3.1.1 Aprendizado Supervisionado

De acordo com FACELLI (2011), no aprendizado supervisionado utilizado para classificação, a meta é encontrar uma hipótese a partir de dados de treinamento que possa prever um valor de que caracterize um novo exemplo, com base nos valores de seus atributos de entrada. As informações de entrada são determinadas pelos padrões do conjunto de dados e possuem um número determinado de atributos, desse modo, assim que for inserido algum outro padrão não rotulado, os algoritmos de aprendizado supervisionado, poderão utilizar padrões adquiridos anteriormente para determinar a qual classe esse novo padrão pertence, de acordo com o conjunto de dados (PARADEDA, 2007).

Algoritmos de AM utilizados no aprendizado supervisionado induzem modelos preditivos, onde um supervisor externo (professor) pode avaliar a capacidade da hipótese induzida de prever o valor de saída para novos exemplos. Conforme mostrado na Figura 7.

Figura 7: Diagrama de blocos da aprendizagem Supervisionada



Fonte: HAYKIN, 2001

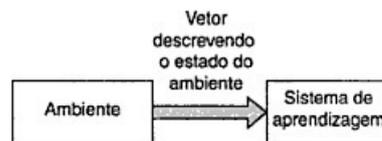
Exemplos de técnicas de aprendizado supervisionado, os quais foram utilizados nesse trabalho são: Árvores de Decisão (AD) (QUILAN; KOHAVI, 1970), Redes Neurais Artificiais

do tipo *Multi-Layer Perceptron* (MLP) (HAYKIN, 2001) e K-Vizinhos mais próximos (K-NN do inglês, *K-Nearest Neighbor* (PATRICK; III, 1970).

3.1.2 Aprendizado Não-Supervisionado

Em técnicas de aprendizado Não-Supervisionado, ao contrário do supervisionado, não existe a presença de um professor externo para supervisionar o processo de aprendizagem, como é mostrado na Figura 8. Essa técnica recebe como entrada padrões sem informação de saída, ou seja, sem classificação e tem como objetivo encontrar padrões distintos em um conjunto de dados, que auxiliem o entendimento desses dados (COSTA, 1999).

Figura 8: Diagrama em blocos da aprendizagem Não-Supervisionada



Fonte: (HAYKIN, 2001)

Técnicas de aprendizado Não-Supervisionado utilizam tarefas de aprendizado descritivo, por meio de agrupamentos ou Clusters, como exemplo dessa técnica pode ser citado o Agrupamento Hierárquico (Sanger, 1989) e K-Médias (do inglês *K-Means*) (PATRICK; III, 1970). Essas técnicas, não serão utilizadas neste trabalho devido às características apresentadas não serem as mais adequadas para o tipo de base que será utilizada.

3.2 Algoritmos Individuais

Nessa seção serão apresentados alguns algoritmos de aprendizado supervisionado, que são os classificadores mais adequados para realizar o aprendizado em uma base biométrica de assinatura. Pois o classificador deve indicar apenas se o padrão (assinatura) apresentado pertence ou não a uma determinada classe (indivíduo).

3.2.1 Algoritmo Árvore de decisão - J48

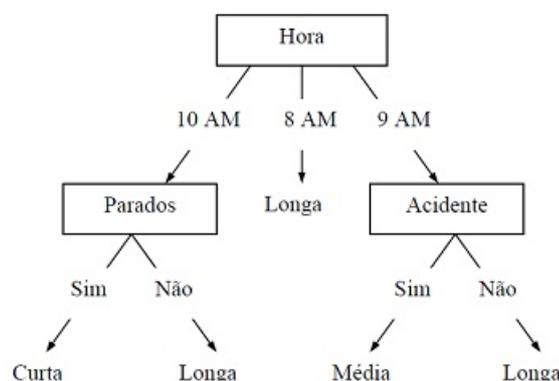
Aprendizado de árvores de decisão é um exemplo de aprendizado indutivo (cria uma hipótese baseada em instâncias particulares que gera conclusões gerais), esse método utiliza a estratégia dividir para conquistar para resolver um determinado problema de

decisão, ou seja, um problema complexo é dividido em problemas mais simples, aos quais é aplicada a mesma estratégia a todos os subproblemas criados (Figura 9). A técnica de AD examina e compara a distribuição de classes durante a construção da árvore e sua estrutura é organizada da seguinte forma:

- cada nó interno (não-folha) é rotulado com o nome de um dos atributos previsores.
- os ramos (ou arestas) saindo de um nó interno são rotulados com valores do atributo naquele nó.
- cada folha é rotulada com uma classe, a qual é a classe prevista para exemplos que pertençam àquele nó folha.

Para melhor compreender o funcionamento de uma árvore de decisão, vamos considerar o seguinte exemplo: Considera-se a situação de ir ao trabalho pela manhã e deseja-se saber qual o tempo estimado da viagem. A Figura 9 apresenta a Árvore de decisão para este problema.

Figura 9: Árvore de decisão para duração de uma viagem



Fonte: (POZZER, 2006)

De acordo com o exemplo mostrado na Figura 9, o padrão da classe será Curta, Média ou Longa. Os valores Hora, Parados e Acidente são atributos do conjunto de dados. Os valores 10, 8 e 9 são valores do atributo Hora, e Sim ou Não são atributos de Parados ou Acidente. Os valores Curta, Média e Longa são os possíveis valores que cada padrão pode receber como rótulo.

As Árvores de decisão são um dos algoritmos mais utilizados tanto em aplicações do mundo real como no meio acadêmico e apresentam inúmeras vantagens, como por exemplo, a flexibilidade, eficiência e seu fácil entendimento e manipulação.

Uma desvantagem apresentada pelas AD é que esse algoritmo é pouco robusto aos dados de grande dimensão, apresentam dificuldade em lidar com atributos contínuos e apresenta grande sensibilidade aos dados de treinamento (MITCHELL, 1997).

Um exemplo de classificador que utiliza a metodologia de árvores de decisão é o algoritmo J48. Esse algoritmo utiliza uma tecnologia chamada *greedy* (Algoritmos Gulosos) para induzir árvores de decisão para posterior classificação e é considerado o mais popular algoritmo do Weka. O J48 constrói um modelo de árvore de decisão baseado num conjunto de dados de treinamento, e usa esse modelo para classificar exatidão do classificador num conjunto de teste. A exatidão do classificador é analisada através da estatística de Kappa, que é uma medida usada em escalas nominais que fornece uma ideia do quanto as observações se afastam daquelas esperadas, fruto do acaso, indicando assim o quão legítimas as interpretações são.

3.2.2 Algoritmo Multi-Layer Perceptron - MLP

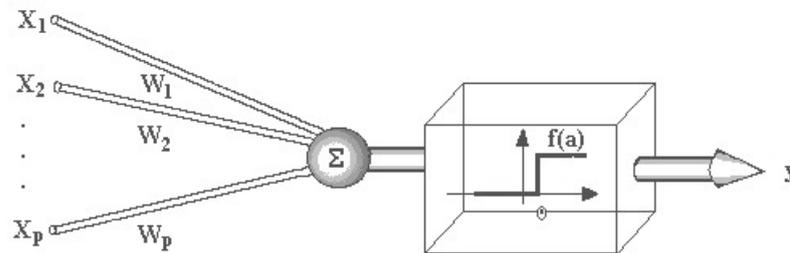
O Algoritmo Multi-Layer Perceptron (MLP) é um classificador de AM bastante utilizado em estudos para problemas no mundo real, sendo um exemplo de RNA. As RNA são técnicas computacionais que apresentam um modelo matemático inspirado no funcionamento dos neurônios como unidades de processamento do cérebro humano e que adquirem conhecimento através da experiência (HAYKIN, 2001).

O primeiro modelo que simula o funcionamento de um neurônio foi proposto por McCullock e Pitts em 1943 (*MCCULLOCK; PITTS, 1943*). De acordo com esse modelo cada neurônio recebe um vetor de entrada com um peso associado, o valor de entrada é multiplicado por esse peso e depois é realizada uma soma ponderada. O número de entradas é igual ao número de atributos ou características que existe no conjunto de dados. Essas entradas são combinadas usando uma função de ativação. Conforme mostrado na Figura 10.

Esses neurônios podem estar dispostos em uma ou mais camadas. Em redes de múltiplas camadas um neurônio recebe em seus terminais de entrada a saída de um neurônio da camada anterior e envia seu valor de saída para uma camada de entrada de um neurônio seguinte (Figura 11).

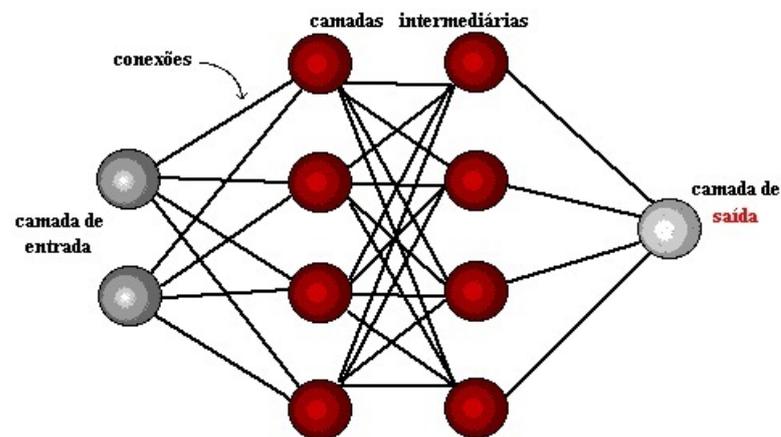
A primeira rede que utilizava o conceito de RNA foi a rede Perceptron, desenvolvida por Rosenblatt (ROSENBLATT, 1958). Essa rede utilizava apenas uma camada de neurônios e os pesos eram ajustados durante o treinamento da rede. É um modelo bastante

Figura 10: Modelo do neurônio proposto por McCulloch e Pitts



Fonte: (FORTES, 2001)

Figura 11: Exemplo de uma RNA multicamadas



Fonte: FORTES, 2001

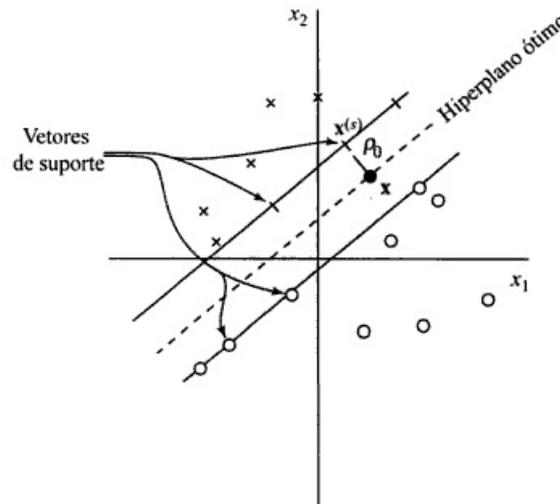
simples, baseado no modelo de neurônio de McCulloch-Pitts, que apresenta uma boa acurácia preditiva em diversos problemas de classificação (FACELI, 2011).

A principal desvantagem da rede Perceptron de uma única camada é que ela só é capaz de aprender problemas linearmente separáveis. Um problema é considerado linearmente separável se for possível separar por um hiperplano os padrões de classes diferentes contidos no espaço de entradas, conforme pode ser visto na Figura 12.

Com o propósito de resolver problemas não-lineares foram criadas as redes com multicamadas, conhecidas como Multi-Layer Perceptron (MLP). Essa rede apresenta uma camada intermediária de neurônios entre as camadas de entrada e saída. As camadas intermediárias utilizam funções de ativação não-lineares, como a sigmoideal.

Mesmo apresentando várias vantagens como tolerância a dados ruidosos, capacidade de representar funções lineares e não-lineares e diversidade de arquiteturas e algoritmos

Figura 12: Exemplo de um hiperplano ótimo para padrões linearmente separáveis



Fonte: HAYKIN, 2001

de treinamentos, as RNA são vistas como uma espécie de caixa preta, pois executam operações difíceis de serem aprendidas e de forma que não permite interpretação a partir de análise visual (FACELLI, 2011).

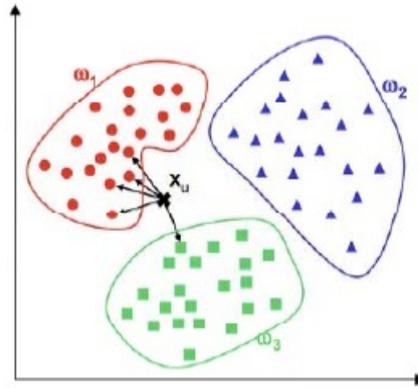
3.2.3 Algoritmo K-Vizinho Mais Próximos - IBK

O algoritmo dos vizinhos mais próximos é um algoritmo baseado em distâncias, onde cada objeto representa um ponto em um espaço definido pelos atributos, denominado espaço de entrada. O algoritmo utiliza uma metodologia que procura pontos semelhantes em meio de pontos diferentes e define uma métrica nesse espaço para que seja possível calcular a distância entre dois pontos, a distância euclidiana é a métrica mais utilizada, mas outras formas como distância Mahalanobis e Hamming Distance também podem ser utilizadas (FACELLI, 2011).

O K-NN (do inglês *K- Nearest Neighbour*) é considerado uma das técnicas de AM mais simples encontradas na literatura (MITCELL, 1997) e possui variações definidas pelo número de vizinhos considerados, sendo a variação 1-NN (1-vizinho mais próximo) a mais simples.

A classificação é realizada verificando em quais classes pertencem os K padrões mais próximos e associando o padrão de teste à classe que for mais frequente entre os K padrões mais próximos, conforme Figura 13.

Figura 13: Exemplo de classificação do método KNN



Fonte: (PROENÇA, 2008)

Não é realizado processamento na fase de treinamento por não ser necessário estimar distribuições de probabilidades das classes. Porém o usuário tem que informar o valor de K-Vizinhos. O desempenho (taxa de acurácia) desse algoritmo depende diretamente do valor de K informado.

Apesar de ser considerado um método simples, facilmente implementado e aplicável mesmo em problemas complexos, o K-NN possui alguns aspectos negativos relacionados à dimensionalidade dos exemplos¹ e intolerância a atributos ruidosos e redundantes (FACELI, 2011).

Um método representativo desse paradigma é o algoritmo IBk, baseado nos conceitos de vizinhança e aprendizado local. O algoritmo IBk (AHA; KIBLER, 1991) é uma implementação da abordagem K-NN classificado como um algoritmo de aprendizado preguiçoso (AHA; KIBLER, 1991), haja vista que não gera nenhum modelo a partir dos dados (utiliza apenas as instâncias em si). Este algoritmo assume que instâncias similares possuam a mesma classificação e dependem fortemente da métrica usada como distância.

3.3 Sistema Multiclassificador

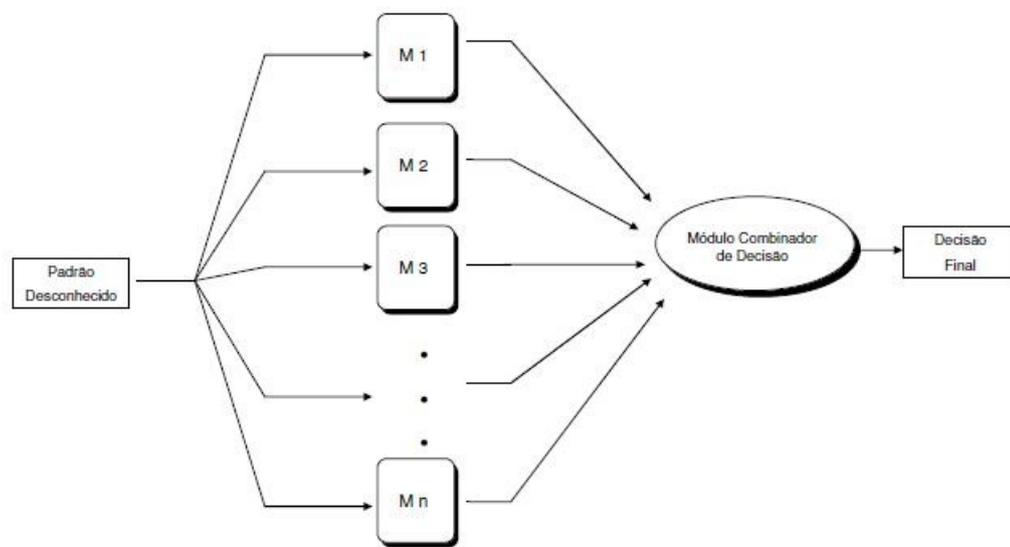
Segundo Cho (1999), um SMC (Sistema Multiclassificador) é composto por vários algoritmos de classificação, e recebe como entrada um conjunto de dados que é processado e depois é gerada uma resposta. Essas respostas são direcionadas a um módulo de combinação, que irá executar uma metodologia de combinação e tomar uma decisão de

¹O número de dimensões ou dimensionalidade é definida pelo número de atributos. Um ponto que está mais perto de outro pode estar muito distante em problemas de alta dimensionalidade.

qual será a melhor solução para o sistema.

A Figura 14 apresenta a estrutura de um SMC, também conhecido como Comitê ou ensemble, onde a segunda camada consiste nos classificadores individuais, que pode ser de arquiteturas iguais ou diferentes, mas com saídas de forma comum e a segunda camada consiste num processo de decisão que recebe como entrada as saídas da camada anterior e realiza uma operação gerando a saída final.

Figura 14: Processo de combinação de decisão de um comitê classificador



Fonte: (VALE, 2009)

Existem vários métodos de combinação de classificadores na literatura, podendo ser agrupados em 2 tipos principais: fusão ² e seleção ³(Santana, 2008).

Neste trabalho serão utilizados os metaclassificadores Soma, Voto e SMO, que são baseados no método de combinação por fusão, a escolha de métodos baseados em fusão foi feita devido ao aproveitamento de todas as informações dos classificadores. Uma breve descrição desses metaclassificadores é feita a seguir:

- Soma: Esse método funciona de maneira que quando apresentado um padrão de entrada, os valores das saídas de cada classificador são somados e a classe vencedora é aquela que possuir o maior valor absoluto (VALE, 2009).
- Voto: A combinação é feita através da votação dos resultados de cada classificador ao ser apresentado um novo exemplo(VALE, 2009).

²utiliza a resposta de todos os classificadores componentes do sistema na construção da resposta final
³onde a resposta final é dada pelo classificador base mais capacitado para o padrão de entrada

- SVM: As SVMs são embasadas pela teoria de aprendizado estatístico. Essa teoria estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu (LORENA; CARVALHO, 2007).

A ideia de utilizar diferentes classificadores em um processo de treinamento surgiu com o objetivo de melhorar o desempenho de um sistema de classificação, pois com o crescimento das pesquisas na área de Classificação de Padrões cada vez mais são testados os limites das técnicas utilizadas para esse princípio. Com isso, está sendo comprovado que a utilização de classificadores individuais têm se tornado pouco eficiente em problemas complexos (FUMERA; ROLI, 2005).

De acordo com Faceli (2011), os Comitês apresentam desempenho superior aos classificadores individuais, pois um agrupamento de diferentes tipos de aprendizado irá apresentar diferentes linguagens de representação, diferentes espaços de procura e diferentes funções de avaliação de hipóteses. Porém, uma condição para que um Comitê apresente desempenho melhor que qualquer um de seus membros é que seus classificadores base devem ser eficientes e diversificados (TAN *et al.*, 2003).

Essa combinação de classificadores pode ser realizada combinando classificadores que utilizam o mesmo algoritmo base (Comitês Homogêneos) ou utilizando algoritmos diferentes (Comitês Heterogêneos), conforme será apresentado nas seções 3.3.1 e 3.3.2:

3.3.1 Combinando Classificadores Homogêneos

Os Comitês homogêneos são métodos que combinam modelos gerados por um único algoritmo. Esse algoritmo é executado várias vezes, utilizando em cada vez uma distribuição diferente de exemplos de treinamento. Essa técnica é utilizada em algoritmos de aprendizagem cuja saída sofre grandes mudanças em resposta a pequenas mudanças nos dados de treinamento (Algoritmos instáveis) (FACELI, 2011). Algumas técnicas baseadas nesse método são Bagging e Boosting e estão apresentadas a seguir. Porém essas técnicas não serão utilizadas neste trabalho.

3.3.1.1 Bootstrap Aggregating (Bagging)

O algoritmo Bagging (do inglês *Bootstrap Aggregating*) produz replicações do conjunto de treinamento por amostragem com reposição, gerando vários subconjuntos diferentes de

treinamento, mas com o mesmo tamanho do original (BREIMAN, 1996). Alguns padrões do conjunto original não aparecem na amostra, enquanto outros podem aparecer mais de uma vez. Para cada subconjunto é gerado um classificador e todos os classificadores são usados para classificar cada exemplo no conjunto de teste e a classificação final é feita geralmente usando um esquema de voto simples (KUNCHEVA, 2004). O Algoritmo 3.1 detalha o funcionamento da técnica Bagging.

ALGORITMO 3.1: Algoritmo de Bagging (PARADEDA, 2007).

- 1: Ler padrões de treinamento X de T (conjunto de treinamento original)
- 2: Selecionar um algoritmo de Aprendizado de Máquina
- 3: Definir o número de iterações k
- 4: para cada iteração $i = 1, 2, \dots, k$ faça
- 5: Formar um subconjunto D_i com o mesmo tamanho do conjunto de treinamento original T
- 6: Gerar o classificador C_i de D_i
- 7: fim para
- 8: Obter a predição final, C^* , através de um esquema de votação de saídas dos k classificadores individuais C_i

A criação de subconjuntos não garante que todos os classificadores induzidos irão generalizar de forma distinta, para que isso ocorra os classificadores devem ser instáveis, no sentido de produzirem comportamentos distintos sempre que forem submetidos a conjuntos de treinamento, como por exemplo, os métodos de Árvores de Decisão e Redes Neurais (BITTENCOURT, 2005).

3.3.1.2 Boosting

Outro algoritmo baseado em técnicas de Comitês homogêneos é o método Boosting que foi proposto por (FREUND; SCHAPIRE, 1996). Esse algoritmo surgiu com a ideia de um conjunto de classificadores fracos originarem um classificador mais forte, onde a ideia de fraco e forte está relacionada a pior ou melhor acurácia (taxa de acertos), respectivamente. O funcionamento desse algoritmo consiste em associar pesos a cada padrão do conjunto de treinamento, os pesos são associados de acordo com a importância do padrão. Um exemplo desse algoritmo é mostrado no Algoritmo 3.2.

ALGORITMO 3.2: Algoritmo de Boosting (PARADEDA, 2007).

- 1: Ler padrões de treinamento X de T (conjunto de treinamento original)
- 2: Selecionar um algoritmo de Aprendizado de Máquina
- 3: Definir o número de iterações k
- 4: Associar um peso $w = 1/n$ para todos os padrões X
- 5: para cada iteração $i = 1, 2, \dots, k$ faça
 - 5: Gerar o classificador C_i considerando os pesos w dos padrões X
 - 6: Atualizar os pesos dos padrões a serem usados pelo próximo classificador
- 7: fim para
- 8: Obter C^* combinando a saída de todos os classificadores C_i por um esquema de votação dos k classificadores individuais C_i .

Uma diferença desse algoritmo em relação ao anterior é que o Bagging atua somente na variância, enquanto o Boosting é capaz de diminuir tanto na variância quanto a taxa de erro de classificação. Isso ocorre devido o fato de aplicar um maior peso aos padrões responsáveis pela queda de desempenho durante o treinamento.

Uma vantagem desse algoritmo é que ele se ajusta a muitos problemas do mundo real, pois os exemplos observados tendem a ter níveis diferentes de dificuldade de classificação.

3.3.2 Combinando Classificadores Heterogêneos

Comitês heterogêneos são métodos que utilizam tipos diferentes de algoritmos de AM na produção dos classificadores. Eles utilizam como dados de treinamento o mesmo conjunto de dados original. Alguns algoritmos que pertencem a essa categoria são o Stacking (*Stacked Generalization*) (Wolpert, 1992) e StackingC (*Stacking with Confidences*) (WOLPERT, 1992).

3.3.2.1 Stackfing

O algoritmo Stacking leva em consideração o fato de certo classificador ser mais seguro que outro (Seewald, 2003), ele usa uma técnica de generalização em pilha, que possui uma arquitetura de aprendizado em camadas. A generalização em pilha é um processo que usa classificadores em camadas mais altas (classificadores de nível 1) para aprender o tipo de erro cometido pelo classificador imediatamente abaixo (classificadores de nível 0) (Faceli, 2011).

O funcionamento do Stacking é descrito em duas fases: na primeira fase (nível 0) um

conjunto de n classificadores base (c_1, c_2, \dots, c_n) é gerado por diferentes algoritmos de aprendizado, utilizando um único conjunto de dados original; na segunda fase (nível 1) o meta classificador combina as saídas geradas em uma predição final, essas informações fornecidas ao meta-classificador um número de atributos igual ao número de classificadores base (PARADEDA, 2007). Um exemplo desse algoritmo é mostrado no Algoritmo 3.3.

ALGORITMO 3.3: Algoritmo de Stacking (PARADEDA, 2007).

```

1: Ler padrões de treinamento  $X$  de  $T$  (conjunto de treinamento original)
2: Selecionar  $n$  algoritmos de Aprendizado de Máquina
3: para cada iteração  $i = 1, 2, \dots, n$  faça
4:   treinar o classificador  $C_i$  com todos os padrões  $X$  e  $T$ 
5:   Obter a saída  $S_i$  do classificador  $C_i$ 
6: fim para
7: Formar um novo conjunto de dados  $D$  de treinamento com os  $S_i$  gerados
8: Treinar o meta-classificador  $M(x)$  com  $D$ 
9: Obter a predição final  $C^*$ 

```

3.3.2.2 StackingC

O algoritmo StackingC é uma variação do Stacking proposta por Seewald (SEEWALD, 2003), onde é realizada uma remoção prévia de características irrelevantes e a redução da dimensionalidade do conjunto de dados usado como entrada para o meta-classificador. Para cada modelo linear associado a uma classe específica, somente a distribuição parcial de probabilidade desta classe é usada no treinamento, diferentemente do Stacking.

Esse método foi motivado devido ao baixo desempenho do Stacking em base de dados com mais de duas camadas. Outro motivo é o fato de que, com poucos atributos, o problema de classificação se torna mais fácil de ser resolvido, já que as informações irrelevantes foram removidas (BITTENCOURT, 2005).

4 *Metodologia dos Experimentos*

Neste capítulo, serão detalhados os procedimentos utilizados para execução dos experimentos e a descrição detalhada do conjunto de dados utilizado para execução dos classificadores e dos comitês, bem como os possíveis resultados.

4.1 Conjunto de dados

Os algoritmos AD, KNN e RNA, descritos no capítulo 3, foram treinados sobre uma base biométrica de assinaturas. No capítulo 2, foi feita uma breve introdução sobre essa base, porém nesta seção serão apresentadas de forma mais detalhada as características da base e de seus atributos.

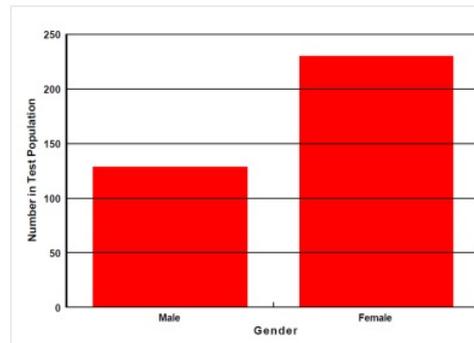
A base biométrica de assinaturas analisada foi coletada pela Universidade de Kent em Canterbury no Reino Unido. As assinaturas foram coletadas no Posto de Correios em Hedge End e no NEWSAGENT em Southampton, ambos localizados no Reino Unido.

As características analisadas de cada participante foram: Nome, Sexo, Faixa Etária (Identificada pelos seguintes agrupamentos: Grupo 1: 18-25, Grupo 2: 26-40, Grupo 3: 41-60, Grupo 4: mais de 60), "Destreza manual"(esquerda, direita, ambidestro), Altura, se possui algum defeito visual e se assinatura mudou recentemente.

Segundo a documentação da base de dados, foram coletadas assinaturas de 359 indivíduos, totalizando 7430 assinaturas, e cada pessoa doou de 2 a 79 assinaturas, dependendo da frequência com que eles visitavam os locais de coleta. Porém, utilizaremos nos testes desse trabalho apenas uma parte da base original contendo 3684 instâncias, pois não foi possível o acesso a base total.

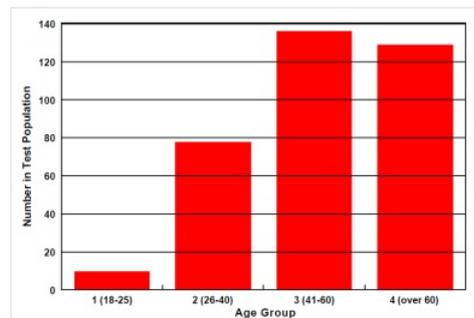
Nas Figura 15 e 16 são apresentados os gráficos com a distribuição de amostras por gênero e distribuição da população estruturada por faixa etária, para que seja possível uma melhor visualização dessas características.

Figura 15: Distribuição da amostra por gêneros



FONTE: Relatório técnico da base de dados de assinatura.

Figura 16: Distribuição da população estudada por faixa etária



FONTE: Relatório técnico da base de dados de assinatura.

Para captura dos dados foi utilizada uma mesa digitalizadora convencional para papel tamanho A4, onde uma simples interface de usuário foi posta na máquina em que a mesa estava conectada, de modo que ficasse invisível para o usuário. A mesa digitalizadora com a caneta era a responsável por toda a interação do sistema.

Cada coleta era iniciada utilizando um cartão magnético que continha a identidade do usuário a ser registrado. O processo de cada assinatura era finalizado ao se apontar com a caneta para um botão OK, contido na mesa digitalizadora. A Figura 17 mostra o modelo do formulário utilizado para a realização das coletas.

4.1.1 Atributos das assinaturas

De acordo com o documento repassado com a base de dados, as amostras são processadas para a extração de diversas características relevantes ao reconhecimento da assinatura. No caso dessa base de assinatura reduzida que será utilizada neste trabalho, tem-se 18 atributos de identificação e mais um atributo classe.

Figura 17: Formulário de coleta

Fonte: Relatório técnico da base de dados de assinaturas

4.1.1.1 Tempo de execução

O período de tempo, em segundos, que o doador utilizou para realizar a assinatura. É calculado com a seguinte fórmula:

$$\text{TempoDeExecuo} = t_{fim} - t_{inicio} \quad (4.1)$$

Onde,

t_{fim} é o último instante em que a caneta esteve em contato com a mesa digitalizadora. t_{inicio} é o primeiro instante em que a caneta esteve em contato com a mesa digitalizadora.

4.1.1.2 Suspensões da caneta

Essa medida guarda a quantidade de vezes em que a caneta perdeu contato com a mesa digitalizadora durante o processo de assinatura, mas excluindo a retirada de finalização. No caso em que a assinatura foi realizada sem nenhuma suspensão, esse atributo receberá o valor 0.

4.1.1.3 Largura da imagem

A largura da imagem obtida em milímetros. Esse atributo é calculado ao subtrair a coordenada x mínima da coordenada x máxima.

$$LarguraDaImagem = MAX(x) - MIN(x) \quad (4.2)$$

4.1.1.4 Altura da imagem

Calculado similarmente ao atributo anterior, mas utilizando o eixo y para o cálculo.

$$AlturaDaImagem = MAX(y) - MIN(y) \quad (4.3)$$

4.1.1.5 Razão entre altura e largura

Esse atributo é a divisão da altura da imagem pela largura, chegando a um valor que mostra as proporções da imagem independentemente do tamanho em si.

$$Razao = \frac{AlturadaImagem}{LarguraDaImagem} \quad (4.4)$$

4.1.1.6 Velocidade horizontal média

Guarda a velocidade média que a caneta percorre no plano x , para isso calcula-se a derivada do deslocamento na coordenada x sobre o tempo. Utilizando o modelo de quatro coeficientes polinomiais de terceira ordem para obter a derivada do deslocamento em cada ponto da coordenada. Utilizando amostras temporais em intervalos constantes, a seguinte aproximação utiliza o valor do deslocamento de quatro conjuntos de coordenadas nos tempos $t - 2$, $t - 1$, $t + 1$ e $t + 2$. As velocidades são calculadas somente enquanto a caneta está em contato com a mesa digitalizadora.

$$\frac{ds}{dt} \approx \frac{1}{12}(-S_{t+2} + 8S_{t+1} + S_{t-2}) \quad (4.5)$$

Onde,

S_{t+2} é a posição do eixo x no tempo $t + 2$.

A velocidade média é obtida somando as velocidades de cada ponto contido no segmento e dividindo pelo número de amostras coletadas.

$$Velocidademedia = \frac{\sum_{t=0}^n \frac{ds}{dt}}{n} \quad (4.6)$$

Onde,

n é o número de amostras.

4.1.1.7 Velocidade vertical média

A velocidade vertical média é calculada utilizando a mesma fórmula da velocidade horizontal média, modificando apenas o eixo x pelo y .

4.1.1.8 Cruzamentos verticais da caneta no ponto central

Esse atributo calcula quantas vezes a caneta passa pelo ponto central no eixo vertical da assinatura. Para isso, são necessários dois passos. Primeiro, a coordenada vertical de todos os pontos da amostra dentro de uma assinatura é calculada. Logo após, é contado quantas vezes a caneta passa por esse ponto ao se escrever a assinatura. A Figura 18 (Assinatura original e projeção do ponto vertical central na assinatura) contém o exemplo de uma assinatura original e da projeção do ponto vertical central na assinatura.

Figura 18: Assinatura original e projeção do ponto vertical central na assinatura



Fonte: Relatório técnico da base de dados de assinaturas

4.1.1.9 Momentos de Invariância

Momentos de invariância podem ser usados para prover informação sobre a forma da imagem. Momentos em imagens binárias podem ser calculados de acordo com a seguinte fórmula.

$$M_{pq} = \sum_{t=1}^N \sum_{j=1}^N t^p j^q f_{tj} \quad (4.7)$$

Onde, N é o tamanho da imagem f_{ij} são valores binários de cada pixel da assinatura (1 = fundo, 0 = imagem).

Dividindo o resultado de imagens de primeira ou maior ordem pelo valor do momento de ordem zero (M00) garante que o momento resultante será invariante pelo tamanho da

imagem. A Tabela 1 mostra as interpretações físicas das dez combinações dos coeficientes de momento que foram extraídas das imagens de cada assinatura. Esses atributos permite diferenciar imagens de assinaturas diferentes.

Tabela 1: Definições dos Momentos

Momento Central	Interpretação física
M00	Número de pontos que compõem a imagem, é a massa da imagem
M10	Soma dos valores das coordenadas horizontais
M01	Soma dos valores das coordenadas verticais
M20	Medida central horizontal
M02	Medida central vertical
M11	Diagonalidade - Indica o quadrante referente ao centróide onde a imagem tem maior massa
M12	Divergência horizontal - Indica a extensão relativa entre a esquerda da imagem comparada com a direita
M21	Divergência vertical - Indica a extensão relativa entre a base da imagem comparada com o topo
M30	Desbalanço horizontal - Local do centro de gravidade da imagem referente a extremidade na metade esquerda
M03	Desbalanço vertical - Local do centro de gravidade da imagem referente a extremidade na metade inferior

4.1.2 Metodologia da Validação

Como metodologia de validação dos resultados foi utilizado um método chamado de Validação Cruzada (do inglês *Cross-Validation*). Este método é utilizado para realizar a execução e comparação dos algoritmos de AM e assegurar que uma medida verdadeira esteja sendo estimada pelos classificadores.

4.1.2.1 Validação Cruzada

Para comparar dois métodos de aprendizado supervisionado é preciso analisar a significância estatística da diferença entre a média da taxa de erro da classificação em conjuntos independentes de teste, dos métodos avaliados (Bittencourt, 2005). O uso do procedimento *K-fold cross validation* (Mitchell, 1997) vem sendo uma boa escolha na avaliação de desempenho de classificadores, nesse procedimento a base de dados é dividida em conjuntos de treinamento e conjunto de validação.

A técnica de Validação Cruzada consiste em dividir a base de dados em k partes (*folds*) mutuamente exclusivas, onde cada uma dessas partições irá conter aproximadamente n/x padrões, sabendo que n é a quantidade de padrões que forma a base. Destas, $k-1$ partes são utilizadas para o treinamento e uma serve como base de testes. O processo é repetido k vezes, de forma que cada parte seja usada uma vez como conjunto de teste. Ao final,

a correção total é calculada pela média dos resultados obtidos em cada etapa, obtendo-se assim uma estimativa da qualidade do modelo de conhecimento gerado e permitindo análises estatísticas (Facelli et al., 2011).

Nesse trabalho, foi utilizada a metodologia de validação cruzada sem estratificação, que é uma variação do procedimento de validação cruzada, que tenta construir partições com as mesmas distribuições de classe não aleatória. Essa metodologia foi escolhida pelo fato da ferramenta utilizada para execução dos algoritmos, a Weka, utilizá-la. Além disso, o valor escolhido para o número de partições (k) foi igual a 10, ou seja, cada execução do treinamento e validação serão utilizados nove partes para treinamento e uma para validação.

4.1.3 Experimentos

Para realização dos experimentos desse trabalho foi escolhida uma base biométrica de assinaturas, como mencionado anteriormente. Considerando o conjunto de dados descrito, o seguinte experimento foi executado: foram executados comitês homogêneos e heterogêneos com três, seis e doze algoritmos. Cada comitê foi executado com uma configuração inicial e com os valores dos atributos modificados em uma média de 20% para mais e para menos (como forma de simplificar iremos chamar de Configuração 1, 2 e 3). Totalizando 90 comitês diferentes.

Nos experimentos realizados, cada algoritmo de AM foi aplicado às diferentes partições obtidas com a utilização da metodologia de validação cruzada, com o número de *folds* igual a 10, ou seja, cada algoritmo foi treinado dez vezes. O treinamento foi realizado utilizando a ferramenta Weka, versão 3.7.7.

No capítulo 5, serão apresentados os resultados dos experimentos realizados e será feita uma análise sobre os melhores resultados de cada metaclassificador. São escolhidos os testes que apresentarem as melhores acurácia e desvio padrão. Os comitês são comparados de modo a verificar em cada caso qual experimento obteve o melhor desempenho, e por fim é destacado o metaclassificador que apresentou melhores taxas de acurácia.

5 *Resultados dos Experimentos*

Neste capítulo são apresentados e interpretados os resultados obtidos através da aplicação dos comitês sobre a base de assinaturas utilizando os algoritmos através da ferramenta WEKA. Primeiramente, são analisados os desempenhos dos multiclassificadores homogêneos e heterogêneos e por último é apresentado um quadro geral com todos os resultados. Os desempenhos serão apresentados e comparados de acordo com o número de algoritmos presentes na formação do comitê e o tipo de metaclassificador, e os melhores resultados são apresentados em negrito.

5.1 Resultados dos Multi-Classificadores Homogêneos

Nesta seção são apresentados os resultados obtidos pelos classificadores homogêneos. Os parâmetros de acurácia e desvio padrão são apresentados para cada resultado obtido de forma individual e através dos comitês. Aborda-se a configuração inicial dos valores dos parâmetros dos comitês, os valores dos parâmetros menores e maiores, e a utilização dos metaclassificadores SMO, IBK, NB, Soma e Voto.

Na Tabela 2, são apresentados a média da taxa de acurácia e do desvio padrão individual e de acordo com os tipos de algoritmos de classificação, ou seja, foi feita uma média e o desvio padrão de todos os resultados obtidos. Pode-se observar através da Tabela que o algoritmo que apresentou melhores resultados em termo de acurácia foi o método de classificação SMO com uma média de 92,94%, 88,75% e 88,42%, para os comitês com três, seis e doze algoritmos de AM, respectivamente.

Na Tabela 3, são mostradas as médias das taxas de acurácia e desvio padrão para os mesmos comitês, porém com os valores de seus parâmetros reduzidos numa média de 20%. Diferentemente da configuração inicial, reduzindo os valores dos parâmetros observa-se que os melhores valores para taxa da média de acurácia foi apresentada pela técnica de classificação soma, apresentando uma média de 95,54%, 95,54% e 96,03%, para

Tabela 2: Média dos comitês homogêneos para o experimento 1

Comitês	Individual	Soma	Voto	Ibk	NB	SMO
3	75,85+-10,00	86,29+-10,7	84,92+-10,68	90,24+-3,03	87,98+-1,27	92,94+-2,4
6	77,32+-11,65	88,79+-10,65	86,96+-9,14	88,00+-4,18	83,75+-3,29	88,75+-6,18
12	75,16+-15,87	87,79+-11,88	86,46+-10,77	86,67+-3,85	84,50+-4,20	88,42+-7,40

os comitês com três, seis e doze algoritmos de AM, respectivamente.

Tabela 3: Média dos comitês homogêneos para o experimento 2

Comitês	Individual	Soma	Voto	Ibk	NB	SMO
3	94,63+-1,11	96,11+-1,05	95,54+-0,94	87,59+-1,84	84,76+-2,05	89,47+-1,59
6	95,02+-1,11	96,28+-0,94	95,54+-0,9	87,68+-1,54	85,36+-1,88	89,59+-1,37
12	94,48+-1,32	96,33+-1,80	96,03+-1,02	87,40+-0,95	83,11+-1,54	89,22+-1,56

Por último, foi analisada a Tabela 4, pode-se observar que nos comitês homogêneos com os valores dos parâmetros modificados em uma média de 20% maior que a configuração original, o método de soma obteve o melhor resultado para taxa de acurácia quando são utilizados três e seis algoritmos na formação do comitê. Nos comitês formados por 12 algoritmos o melhor desempenho pode ser observado utilizando o meta-classificador SMO, que também apresenta a melhor taxa de desvio padrão. As melhores médias de acurácia foram 95,54%, 95,79% e 89,22%, para os comitês com três e doze algoritmos de AM, respectivamente.

Tabela 4: Média dos comitês homogêneos para o experimento 3

Comitês	Individual	Soma	Voto	Ibk	NB	SMO
3	77,58+-6,18	95,54+-0,68	93,43+-0,55	87,09+-1,85	75,39+-2,86	89,47+-1,59
6	76,49+-6,59	95,79+-0,87	95,11+-1,01	83,66+-1,89	76,25+-3,39	89,59+-1,37
12	50,95+-11,43	80,62+-9,7	68,49+-14,01	81,25+-4,55	69,66+-5,49	89,22+-1,56

De acordo com os resultados obtidos e apresentados nas Tabelas 2, 3 e 4, é possível perceber que para os comitês homogêneos o melhor resultado para taxa de acurácia foi obtido pelos comitês constituídos de 12 algoritmos de AM utilizando a técnica de voto como metaclassificador, onde foi obtida uma taxa de 96,33%.

Observando os desvios padrão, percebe-se que quando houve alteração nos valores dos parâmetros iniciais houve uma melhora significativa, conforme pode ser observado na Tabela 3.

5.2 Resultados dos Multiclassificadores Heterogêneos

Seguindo o mesmo padrão de análise da seção anterior, nesta seção são analisadas as médias para taxa de acurácia e desvio padrão para os comitês heterogêneos.

Na tabela 5, são apresentados os resultados para a configuração inicial dos valores dos parâmetros. Para essa configuração o metaclassificador SMO obteve os melhores resultados quando utilizado como metaclassificador. Alcançando médias de 86,98%, 89,75% e 90,48% para os comitês com 3, 6 e 12 algoritmos, respectivamente.

Tabela 5: Média dos comitês heterogêneos para o experimento 1

Comitês	Individual	Soma	Voto	Ibk	NB	SMO
3	78,18+-10,03	89,39+-5,63	87,04+-7,75	86,01+-3,68	77,33+-4,44	86,98+-3,08
6	75,94+-6,7	89,62+-8,35	88,09+-7,34	88,98+-3,27	85,33+-3,45	89,75+-2,85
12	76,69+-7,09	90,23+-6,87	88,86+-7,08	89,69+-2,99	87,17+-3,29	90,48+-2,61

Quando foram reduzidos os valores dos parâmetros para esses comitês, observa-se na Tabela 6, que o algoritmo SMO continua apresentando os melhores resultados e que esses resultados foram bastante melhorados. 92,93%, 89,75% e 90,48%, para os comitês com 3, 6 e 12 algoritmos, respectivamente.

Tabela 6: Média dos comitês heterogêneos para o experimento 2

Comitês	Individual	Soma	Voto	Ibk	NB	SMO
3	89,45+-1,501	90,28+-1,55	90,14+-1,48	90,24+-1,51	87,97+-1,69	92,93+-1,23
6	89,07+-1,64	90,58+-1,48	89,94+-1,47	90,97+-1,57	88,91+-1,76	93,66+-1,12
12	88,15+-1,66	90,77+-1,54	90,94+-1,59	91,33+-1,60	88,93+-1,51	93,49+-1,36

Tabela 7: Média dos comitês heterogêneos para o experimento 3

Comitês	Individual	Soma	Voto	Ibk	NB	SMO
3	82,64+-4,98	88,96+-2,33	83,56+-4,37	88,78+-1,84	81,81+-3,05	91,55+-1,23
6	86,62+-3,46	90,25+-1,57	88,67+-1,90	90,12+-1,56	82,35+-3,98	91,15+-1,12
12	85,05+-3,08	89,7+-1,83	88,52+-1,89	90,45+-1,55	86,33+-2,06	92,78+-1,36

Na Tabela 7, mesmo aumentando o valor dos parâmetros ainda se obtêm os melhores valores para taxa de aprendizado através da utilização do metaclassificador SMO. Porém percebe-se que com o aumento houve uma queda no aprendizado. Podemos observar

também através das tabelas que foram obtidos bons resultados para média da taxa de desvio padrão, onde na maioria dos casos que foram obtidos as melhores taxas de acerto, também foram observados as melhores taxas de desvio padrão.

5.3 Melhores resultados dos MultiClassificadores

Nesta seção são apresentados e analisados os valores obtidos para todos os comitês utilizados nesse trabalho. São analisados os valores de cada iteração, a média da taxa de acurácia e o desvio padrão para cada comitê e cada metaclassificador. Por motivo de simplicidade, serão comparados os comitês para os três experimentos realizados de acordo com o algoritmo metaclassificador.

Analisando as tabelas 5, 6 e 7 podemos observar que em quase todos os casos os melhores resultados para taxa de acurácia foi com a utilização do metaclassificador SMO. Somente quando são utilizados 3 algoritmos de AM no experimento 1 é que temos um resultado melhor utilizando a técnica por soma.

Podemos observar também através das tabelas que foram obtidos bons resultados para média da taxa de desvio padrão, onde na maioria dos casos que foram obtidos as melhores taxas de acerto, também foram observados as melhores taxas de desvio padrão.

5.3.1 Resultados do Metaclassificador SMO

Nesta subseção são mostrados os resultados do metaclassificador SMO aplicado sobre os comitês utilizados e para as três configurações a qual foram testados. Na tabela 8, são destacados em negrito os melhores resultados da execução do metaclassificador SMO sobre cada comitês. Analisando a tabela os melhores resultados para taxa de acerto, nos três experimentos, foram apresentados pelos algoritmos que utilizam redes neurais em sua composição, em todo esse trabalho é utilizado o algoritmo MLP (*Multi-Layer Perceptron*). Percebe-se também que o número de redes neurais utilizadas não é um fator determinante para a melhora do desempenho, visto que mesmo aumentando o número de redes a média da taxa de acurácia em alguns casos não aumentou.

Tabela 8: Resultados do Metaclassificador SMO (*Sequential Minimal Optimization*)

Comitês	Experimento 1	Experimento 2	Experimento 3
1N1A1K	91,13+-1,89	93,82+-1,47	94,94+-0,98
2A1K	77,38+-3,89	88,65+-1,51	90,49+-1,60
2A1N	77,38+-3,89	92,49+-0,87	93,48+-1,28
2K1A	89,75+-2,55	91,07+-1,39	91,75+-1,52
2K1N	89,63+-2,72	94,16+-1,32	95,23+-1,09
2N1A	91,88+-3,6	95,14+-1,02	84,75+-3,65
2N1K	91,75+-3,06	95,24+-1,09	90,25+-2,74
3AD	74,5+-4,98	81,33+-2,36	81,49+-2,71
3KNN	90,5+-3,07	91,24+-1,28	91,61+-1,66
3NN	92,38+-3,07	95,85+-1,14	93,04+-2,24
2N2A2K	93,88+-2,53	94,87+-1,03	94,67+-0,88
4A2K	84,13+-4,45	90,05+-1,82	90,76+-1,70
4A2N	76,88+-3,64	93,68+-0,77	92,77+-1,16
4K2A	93,25+-2,31	91,34+-1,35	92,02+-1,69
4K2N	89,75+-2,25	94,63+-1,1	93,68+-0,98
4N2A	95,25+-3,01	95,31+-0,84	88,30+-10,48
4N2K	95,13+-1,81	95,75+-1,00	85,92+-12,55
6AD	81,75+-5,92	81,39+-2,15	81,63+-2,84
6KNN	91,00+-2,73	91,14+-1,06	91,44+-1,61
6NN	93,50+-2,83	96,26+-0,91	94,21+-9,23
4N4A4K	94,50+-2,51	94,66+-1,46	95,65+-1,1
8A4K	85,63+-3,81	90,08+-1,89	90,49+-1,91
8A4N	80,75+-4,3	93,58+-1,43	93,31+-0,89
8K4A	91,5+-2,14	91,03+-1,33	92,15+-1,48
8K4N	88,88+-2,85	93,89+-1,22	92,32+-1,61
8N4A	96,25+-1,16	95,52+-1,2	94,58+-6,18
8N4K	95,88+-1,55	95,68+-1,02	94,90+-4,76
12AD	80,13+-6,13	81,16+-2,15	81,50+-2,43
12KNN	90,75+-2,71	90,90+-1,35	91,47+-1,4
12NN	94,38+-2,2	95,62+-1,2	71,32+-19,17

De acordo com os resultados obtidos, percebe-se que há diferenças significativas em alguns resultados. Um exemplo claro é o uso de 12 RNA, onde temos ótimas taxas de aprendizado para as duas primeiras configurações dos valores dos parâmetros (94,38% e 95,62%), mas quando há um aumento nesses valores a taxa de acurácia diminui consideravelmente (71,32%). Uma explicação para essa queda bastante significativa é o aumento dos valores dos parâmetros, principalmente na taxa de aprendizagem, visto que ao passo que se é colocada uma taxa de aprendizagem maior significa uma aprendizagem mais rápida, porém começará a apresentar picos de aprendizagem, ou seja, há uma memorização dos resultados e não aprendizado. A melhor taxa de aprendizado foi obtida pelo comitê

6NN, composto por seis redes neurais, e com o número de parâmetros reduzidos. Onde conseguiu-se a taxa de 96,26% de aprendizado, mas houveram outros exemplos que obtiveram taxas de aprendizado bastante próximas a essa, como é o caso do comitê 8N4A que apresentou 96,25% de taxa de aprendizado.

5.3.2 Resultados do Metaclassificador Ibk

A Tabela 9, mostra os resultados da execução do metaclassificador Ibk sobre os comitês. Os valores em negrito indicam os melhores valores encontrados para taxa de acurácia e desvio padrão, seguindo o mesmo padrão de análise da seção anterior.

Observa-se que utilizando esse metaclassificador, assim como o SMO, os melhores resultados são obtidos pelos comitês que utilizam o MLP em sua composição, outra semelhança é o fato de não existir dependência do resultado com o número de redes neurais utilizadas na formação do comitê. Porém, diferentemente do classificador SMO, utilizando esse meta-classificador os melhores resultados são apresentados pelos comitês com a configuração 1 e 3. O melhor resultado foi obtido na utilização do comitê 8N4K, que apresentou 96,75% de taxa de aprendizado, para a configuração inicial dos parâmetros.

5.3.3 Resultados do Metaclassificador NB

Nesta subseção são analisados os resultados da utilização do metaclassificador Naive Bayes. A Tabela 10 apresenta todos os resultados desse metaclassificador sobre os algoritmos estudados neste trabalho, onde estão novamente destacados em negrito o melhor resultado para cada cada comitê.

Analisando a tabela, percebe-se que houve uma diminuição da taxa de acurácia, em relação aos métodos das subseções anteriores. A maioria dos exemplos a taxa de acurácia não alcança o índice de 90%. Na última linha da tabela, onde utilizamos 12 RNA foi obtida uma média de apenas 57,23%, muito inferior as médias que alcançamos utilizando os outros métodos de classificação para taxa de acurácia, sendo o pior valor encontrado nas implementações realizadas neste trabalho. A explicação para essa queda no aprendizado deve-se ao fato do algoritmo Naive Bayes trabalhar melhor com atributos irrelevantes¹ e quando os atributos são independentes das classes dadas.

Mesmo com essa queda na taxa de aprendizado, os melhores resultados para taxa de acurácia são apresentados pelos comitês que utilizam o MLP, porém não há grandes

¹Um atributo é irrelevante se ele não é capaz de distinguir exemplos de uma mesma classe

Tabela 9: Resultados do Metaclassificador Ibk

Comitês	Experimento 1	Experimento 2	Experimento 3
1N1A1K	88,88+-3,76	91,65+-1,23	93,41+-1,5
2A1K	73,38+-4,53	85,04+-1,78	85,87+-1,94
2A1N	73,38+-4,53	87,23+-1,68	87,78+-1,88
2K1A	90,38+-3,62	90,52+-1,33	88,62+-1,7
2K1N	90,38+-3,62	91,2+-1,53	93,21+-1,25
2N1A	92,88+-2,7	92,59+-1,59	80,95+-3,14
2N1K	92,88+-3,04	93,48+-1,49	91,64+-1,49
3AD	69,63+-5,50	79,36+-2,84	78,44+-2,86
3KNN	88,63+-3,74	88,96+-1,42	88,62+-1,92
3NN	93,00+-3,21	94,46+-1,27	94,23+-0,77
2N2A2K	94,75+-1,04	92,50+-1,56	93,07+-1,08
4A2K	78,38+-6,14	85,60+-1,75	86,28+-2,16
4A2N	75,63+-6,02	88,22+-1,71	82,85+-2,65
4K2A	92,75+-2,92	89,37+-2,02	91,67+-2,18
4K2N	89,63+-2,56	92,46+-1,47	92,12+-1,52
4N2A	96,13+-2,03	93,65+-0,87	92,97+-0,99
4N2K	95,63+-2,20	95,04+-1,66	94,18+-0,52
6AD	76,16+-6,85	78,54+-2,17	78,81+-2,20
6KNN	93,75+-2,55	89,03+-1,38	88,52+-1,60
6NN	94,13+-3,14	95,48+-1,08	91,67+-2,18
4N4A4K	95,50+-1,31	93,68+-1,44	94,50+-1,06
8A4K	83,63+-5,76	86,28+-1,72	86,76+-2,14
8A4N	77,63+-4,96	88,56+-1,994	89,44+-1,47
8K4A	89,88+-3,36	89,81+-1,51	90,49+-1,58
8K4N	88,00+-3,25	92,39+-1,41	91,07+-1,50
8N4A	96,50+-1,41	93,92+-1,48	91,87+-1,09
8N4K	96,75+-0,89	94,70+-1,76	92,03+-1,43
12AD	74,88+-6,33	78,30+-1,96	78,62+-2,00
12KNN	90,13+-2,95	88,90+-1,20	88,96+-2,48
12NN	95,13+-2,30	95,01+-1,70	76,17+-9,20

diferenças em relação aos multiclassificadores sem redes neurais.

5.3.4 Resultados do Metaclassificador Soma

Nesta subseção são analisados os resultados para média da taxa de acurácia e desvio padrão quando é utilizada a técnica de classificação soma. Para análise da tabela é seguido o mesmo padrão das análises anteriores. Observa-se que de modo geral o método de soma apresenta resultados melhores quando os valores dos parâmetros foram reduzidos, porém o melhor aprendizado apresentado por esse método foi obtido pelo comitê 8N4k com a

Tabela 10: Resultados do Metaclassificador Naive Bayes

Comitês	Experimento 1	Experimento 2	Experimento 3
1N1A1K	86,50+-3,63	89,06+-1,47	89,51+-1,07
2A1K	55,25+-6,45	86,42+-1,97	84,89+-2,08
2A1N	55,25+-6,45	86,45+-1,98	83,26+-1,77
2K1A	87,88+-2,8	89,58+-1,63	89,64+-1,96
2K1N	86,88+-4,12	88,93+-1,42	90,25+-0,92
2N1A	85,25+-3,81	87,47+-1,78	54,11+-10,24
2N1K	84,38+-3,85	87,95+-1,59	81,05+-3,34
3AD	57,75+-5,44	78,26+-2,26	60,89+-1,95
3KNN	88,25+-2,96	88,39+-1,95	87,84+-1,98
3NN	82,38+-3,07	87,91+-1,95	77,46+-4,68
2N2A2K	84,88+-3,18	90,56+-1,70	84,65+-3,33
4A2K	84,63+-4,69	84,79+-2,16	81,62+-1,83
4A2N	72,88+-4,19	88,62+-2,07	78,92+-3,01
4K2A	87,25+-3,24	90,19+-1,43	88,73+-2,66
4K2N	89,63+-2,56	92,46+-1,47	92,12+-1,52
4N2A	91,63+-2,92	88,01+-1,56	80,14+-4,31
4N2K	94,13+-2,70	89,34+-1,39	79,36+-7,47
6AD	73,13+-4,02	77,05+-2,62	64,34+-2,19
6KNN	88,25+-3,69	90,25+-1,49	89,03+-1,54
6NN	89,88+-2,17	88,79+-1,53	75,38+-6,44
4N4A4K	93,13+-2,64	90,49+-1,33	89,85+-2,56
8A4K	78,38+-5,40	84,65+-1,96	82,14+-2,08
8A4N	74,00+-5,24	86,08+-1,40	81,40+-2,65
8K4A	90,75+-2,76	88,66+-1,47	87,23+-1,52
8K4N	89,75+-2,19	90,80+-1,29	91,07+-1,50
8N4A	92,25+-2,92	90,73+-1,29	84,43+-2,01
8N4K	92,00+-1,93	91,13+-1,54	85,59+-1,32
12AD	72,63+-6,99	70,67+-1,70	64,66+-2,83
12KNN	88,50+-2,33	88,28+-1,54	87,10+-1,52
12NN	92,38+-3,29	90,01+-1,39	57,23+-12,12

configuração inicial dos parâmetros, alcançando uma taxa de acurácia de 96,63%.

5.3.5 Resultados do Metaclassificador Voto

Nesta subseção são apresentados os resultados da utilização do método de classificação por voto. As análises são realizadas seguindo o padrão das análises anteriores.

Para esse metaclassificador pode-se observar através da Tabela 12 que para a maioria dos comitês foram obtidos melhores resultados quando os valores dos parâmetros foram

Tabela 11: Resultados do Metaclassificador Soma

Comitês	Experimento 1	Experimento 2	Experimento 3
1N1A1K	90,29+-2,29	93,91+-1,5	94,89+-0,9
2A1K	81,38+-6,50	81,78+-2,32	83,68+-2,47
2A1N	81,38+-6,50	81,78+-2,32	84,2+-2,46
2K1A	92,25+-2,96	92,34+-1,34	92,10+-1,48
2K1N	92,25+-2,96	92,61+-1,25	94,4+-1,15
2N1A	94,13+-3,04	94,78+-0,9	86,72+-3,9
2N1K	94,13+-3,04	94,78+-1,28	86,78+-3,96
3AD	74,13+-6,77	81,75+-2,27	80,99+-2,55
3KNN	90,50+-2,93	91,31+-1,13	91,5+-1,5
3NN	94,25+-2,87	96,11+-1,05	95,54+-0,68
2N2A2K	95,25+-1,83	94,19+-1,3	94,10+-1,34
4A2K	77,50+-6,00	81,73+-2,24	83,46+-2,41
4A2N	77,63+-6,05	81,73+-2,24	83,63+-2,45
4K2A	93,13+-2,10	92,58+-1,23	92,37+-1,85
4K2N	92,25+-1,83	92,67+-1,42	93,67+-1,25
4N2A	96,25+-2,25	95,27+-1,06	93,75+-1,64
4N2K	95,38+-2,72	95,9+-0,89	95,11+-0,97
6AD	76,50+-5,83	81,73+-2,24	81,73+-2,61
6KNN	94,75+-1,75	91,28+-1,12	91,33+-1,58
6NN	95,13+-2,90	96,28+-0,94	95,79+-0,87
4N4A4K	95,63+-1,85	94,46+-1,24	94,67+-1,5
8A4K	81,50+-4,96	83,27+-2,38	84,03+-2,43
8A4N	80,13+-4,29	83,35+-2,45	84,36+-2,29
8K4A	90,50+-2,67	91,61+-1,34	92,53+-1,72
8K4N	91,13+-2,23	91,31+-1,23	92,91+-1,22
8N4A	96,13+-2,30	95,57+-0,94	95,89+-2,03
8N4K	96,63+-1,77	95,84+-1,22	96,01+-1,87
12AD	74,13+-5,17	82,11+-2,5	81,40+-2,35
12KNN	93,63+-2,13	91,28+-1,11	91,69+-1,64
12NN	95,63+-2,00	96,33+-1,39	80,62+-9,7

modificados, principalmente para configuração 2.

O melhor resultado foi obtido pelo comitê 12RNA, com os menores valores dos parâmetros, alcançando 96,03% de taxa de acurácia. A medida que foi aumentado os valores para os parâmetros esse mesmo comitê atingiu o pior resultado para esse metaclassificador, apresentando uma taxa de aprendizado de 68,49%. A explicação para essa grande diferença nos resultados é a mesma exposta na subseção 5.3.1.

Tabela 12: Resultados do Metaclassificador Voto

Comitês	Experimento 1	Experimento 2	Experimento 3
1N1A1K	92,13+-2,47	93,48+-1,40	93,29+-1,41
2A1K	75,75+-4,59	81,78+-2,32	83,57+-2,59
2A1N	75,75+-4,59	81,78+-2,32	84,46+-2,79
2K1A	90,63+-3,25	91,44+-1,33	92,36+-1,49
2K1N	90,25+-3,45	92,58+-1,25	92,53+-1,29
2N1A	92,38+-2,83	94,87+-0,77	69,17+-10,97
2N1K	92,38+-2,83	95,03+-1,03	72,53+-10,11
3AD	72,75+-5,97	81,73+-2,24	81,02+-2,79
3KNN	89,25+-3,69	91,69+-1,40	90,87+-1,94
3NN	92,75+-3,06	95,54+-0,94	93,43+-0,55
2N2A2K	94,13+-1,64	93,18+-1,58	92,53+-1,68
4A2K	77,88+-4,94	81,70+-2,27	81,59+-2,85
4A2N	77,63+-5,34	81,73+-2,24	81,56+-2,45
4K2A	89,88+-3,09	91,06+-1,27	91,01+-1,98
4K2N	89,25+-2,96	92,56+-1,19	91,42+-1,73
4N2A	96,25+-2,25	95,27+-1,06	93,75+-1,94
4N2K	93,63+-3,11	94,92+-0,77	93,26+-1,26
6AD	76,63+-5,10	81,70+-2,20	81,40+-2,77
6KNN	90,25+-3,15	90,71+-1,68	89,68+-2,27
6NN	94,00+-2,98	95,54+-0,90	95,11+-1,01
4N4A4K	95,13+-1,13	94,08+-1,25	93,75+-1,53
8A4K	79,75+-5,34	83,25+-2,48	82,54+-2,39
8A4N	78,00+-4,38	83,49+-2,50	82,35+-2,50
8K4A	90,50+-2,33	91,82+-1,49	91,93+-1,58
8K4N	90,25+-2,49	92,62+-1,32	92,01+-1,48
8N4A	94,25+-2,76	95,54+-0,96	93,50+-1,53
8N4K	94,13+-2,70	95,76+-1,18	94,23+-2,32
12AD	74,13+-4,61	81,94+-2,72	81,05+-2,49
12KNN	91,25+-2,49	91,61+-1,46	91,28+-1,52
12NN	94,00+-2,67	96,03+-1,02	68,49+-14,10

5.3.6 Resultados dos Comitês rodando individualmente

Nesta subseção são analisados os resultados para média da taxa de acurácia e desvio padrão individual. Para análise da tabela é seguido o mesmo padrão das análises anteriores. Observa-se que de modo geral o método de soma apresenta os melhores resultados e que na maioria dos exemplos a taxa de acurácia com a utilização de algum meta classificador é melhor que o resultado individual. Percebe-se também que na maior parte dos exemplos os comitês que possuem o algoritmo Ibk obtiveram os melhores resultados. Somente para configuração com os menores valores de parâmetros é que os melhores re-

sultados são observados em comitês com a presença do algoritmo MLP, contudo nota-se que utilizando menores taxas para os parâmetros foram obtidos melhores resultados de aprendizagem, alcançando a taxa de 95,02% de acurácia para o comitê 6RNA.

Tabela 13: Resultados individual

Comitês	Experimento 1	Experimento 2	Experimento 3
1N1A1K	81,67+-3,24	89,33+-1,40	89,1+-7,8
2A1K	63,79+-3,92	86,12+-1,92	84,79+-2,00
2A1N	63,79+-3,92	86,12+-1,92	85,73+-2,2
2K1A	86,29+-2,29	87,92+-1,55	87,78+-1,78
2K1N	86,88+-2,30	92,57+-1,15	92,34+-1,39
2N1A	82,42+-2,48	90,51+-1,49	67,71+-10,04
2N1K	82,42+-2,48	93,60+-1,12	71,06+-9,68
3AD	64,33+-4,19	81,6+-2,53	80,88+-2,70
3KNN	90,79+-2,54	91,06+-1,38	90,25+-1,61
3NN	82,42+-2,30	94,63+-1,11	77,58+-6,18
2N2A2K	74,65+-2,75	89,11+-1,6	84,06+-3,17
4A2K	66,94+-3,43	84,76+-2,02	84,28+-2,37
4A2N	67,06+-3,47	86+-2,04	80,94+-3,78
4K2A	79,35+-2,01	87,63+-1,68	87,42+-2,10
4K2N	78,25+-1,81	92,28+-1,29	87,29+-3,08
4N2A	83,29+-2,37	90,21+-1,60	76,55+-5,4
4N2K	82,06+-2,74	93,50+-1,28	77,83+-4,38
6AD	64,02+-3,11	81,55+-2,46	80,90+-2,75
6KNN	85,71+-2,18	90,69+-1,52	90,00+-1,88
6NN	82,23+-2,25	95,02+-1,11	76,49+-6,59
4N4A4K	74,66+-2,60	89,06+-1,66	87,00+-3,05
8A4K	69,93+-3,36	84,57+-2,14	83,96+-2,28
8A4N	66,11+-3,29	85,81+-2,10	82,16+-3,15
8K4A	78,70+-1,87	81,71+-1,69	87,49+-1,99
8K4N	77,97+-2,14	92,11+-1,29	84,65+-4,93
8N4A	83,20+-2,06	90,34+-1,58	81,54+-3,98
8N4K	86,28+-2,34	93,42+-1,19	82,76+-2,46
12AD	56,89+-2,97	81,25+-2,54	80,40+-2,51
12KNN	85,42+-2,35	90,99+-1,41	90,68+-1,56
12NN	83,19+-1,76	94,48+-1,32	50,95+-11,43

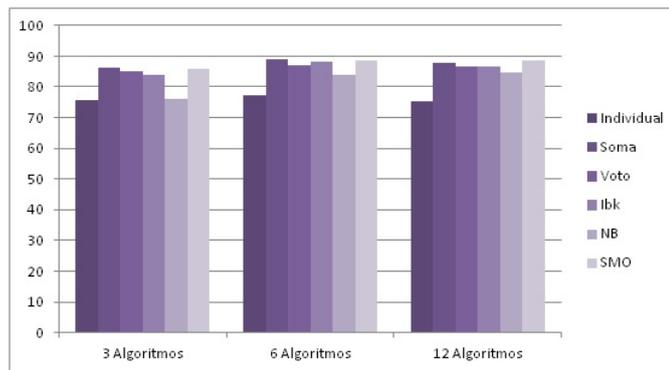
5.3.7 Discussão dos resultados

Por questão de praticidade e para melhor visualização das tabelas apresentadas nas seções 5.1 e 5.2, onde são apresentados os resultados das médias das taxas de acurácia e de desvio padrão, são criados os gráficos a seguir. A Figura 18, mostra um gráfico onde

se é possível visualizar para cada metaclassificador utilizado nas implementações desse trabalho o desempenhos dos comitês homogêneos para a configuração 1.

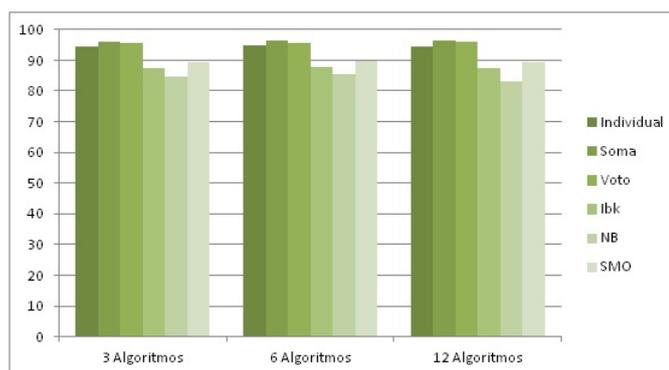
Observa-se, que as taxas de acurácia variam entre pouco mais de 70% a quase os 90%. Tendo um desempenho levemente melhor quando utilizamos comitês com 6 e 12 algoritmos.

Figura 19: Média dos comitês homogêneos para configuração inicial dos valores dos parâmetros.



Quando foram reduzidos os valores dos parâmetros (experimento 2), houve uma melhora significativa na média de aprendizado para todos os classificadores. Apresentando uma variação 83% a 96% de taxa de acurácia.

Figura 20: Média dos comitês homogêneos para configuração com valores dos parâmetros menores.



Após a análise da Figura 18, Figura 19 e Figura 20, que apresentam os gráficos para as médias de acurácia dos multiclassificadores homogêneos, pode-se observar que os melhores resultados foram obtidos pelo classificador soma. Através de uma análise mais detalhada de todos os resultados apresentados nas tabelas da seção 5.3 observa-se que os melhores resultados para todos os metaclassificadores são obtidos pelos comitês que utilizam o metaclassificador MLP, porém para a base de dados e os atributos analisados

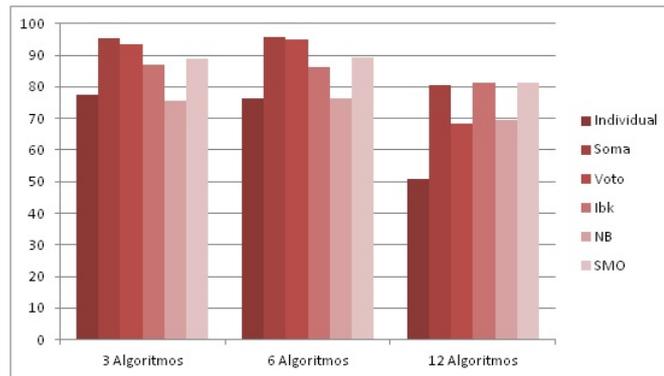


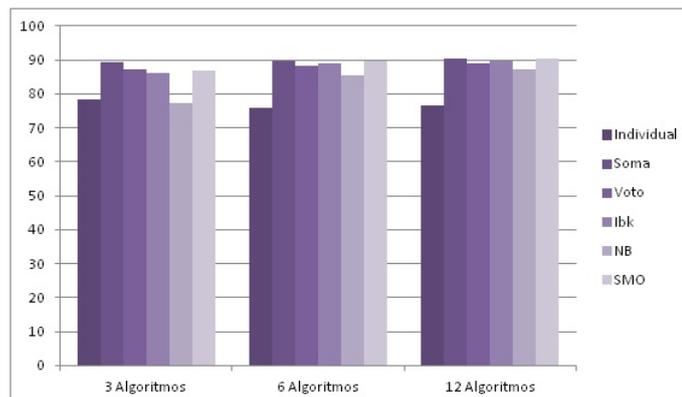
Figura 21: Média dos comitês homogêneos para configuração com valores dos parâmetros maiores.

a quantidade de redes neurais que compõe o comitê não é importante, ou seja, mesmo aumentando a quantidade de redes neurais não há melhora na taxa de aprendizado. Esse resultado deve-se ao fato de o MLP ser um algoritmo mais robusto que o IBK e o NB e lidar melhor com conjuntos de dados que apresentam um grande número de classes e atributos numéricos.

O algoritmo NB não teve resultados bons devido o fato desse algoritmo apresentar bons resultados quando se tem conjuntos de dados com muitos atributos, ou seja, quando tem-se atributos irrelevantes. Sendo assim, é explicado o resultado não tão bom para esse algoritmo para a base estudada neste trabalho, onde contém apenas 18 atributos.

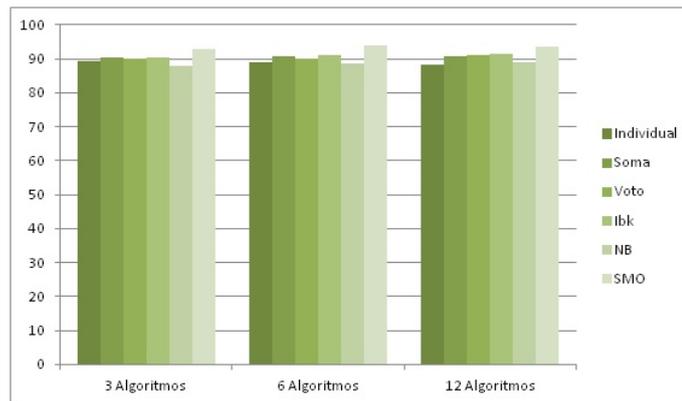
Analisando a Figura 21, que apresenta o gráfico onde são apresentados os resultados para os comitês heterogêneos com a configuração inicial para os valores dos atributos, houve um aumento nas taxas de acurácia quando modificamos os valores dos atributos e utilizamos um metaclassificador, porém a taxa individual teve uma leve queda.

Figura 22: Média dos comitês heterogêneos para configuração inicial dos valores dos atributos.



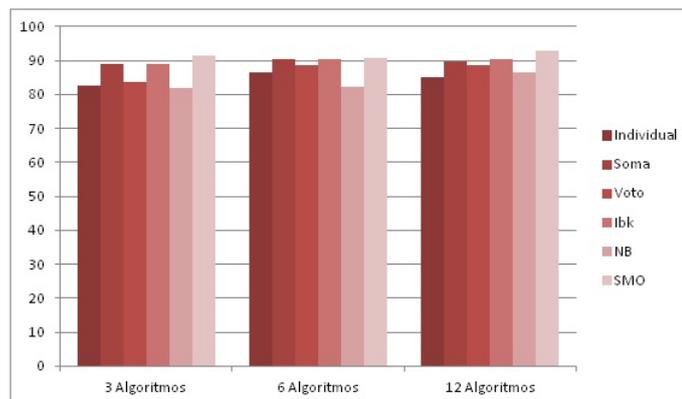
Na Tabela 22, pode-se observar que assim como os comitês homogêneos, os heterogêneos também apresentaram melhores taxas de acurácia quando os valores dos atributos foram modificados, na configuração 2. Alcançando médias por volta dos 90%.

Figura 23: Média dos comitês heterogêneos para configuração com valores dos atributos menores.



Quando houve um aumento no valor dos atributos (configuração 3), é possível notar que não há mais uma homogeneidade nos resultados dos classificadores. Na figura 23, pode-se observar que tivemos redução em alguns classificadores, alcançando uma média de aproximadamente 80%. Os classificadores voto, soma e SMO continuaram com uma média de aprendizado por volta dos 90%.

Figura 24: Média dos comitês heterogêneos para configuração com valores dos atributos maiores.



6 *Conclusões e trabalhos futuros*

Após todas as análises pode-se concluir sobre:

- Desempenho em relação ao conjunto de dados com o número de parâmetros: observou-se que o desempenho dos comitês utilizando menor valor para os números de parâmetros obteve os melhores resultados, verificando de forma generalizada.
- Desempenho em relação à estrutura dos comitês: percebeu-se que os comitês homogêneos obtiveram melhor desempenho que os heterogêneos quando foi modificado os valores dos parâmetros.
- Desempenho em relação ao método de combinação: observou-se que para os comitês homogêneos os métodos mais simples como soma e voto alcançaram os melhores valores para taxa de acurácia, porém quando é realizada a análise dos comitês heterogêneos observa-se que há uma leve diferença entre os resultados, mas quem obtém os melhores resultados é o método SMO.

6.1 **Trabalhos Futuros**

Alguns trabalhos futuros podem ser desenvolvidos a partir do estudo realizado neste trabalho, como forma de expandir as análises realizadas. Conforme listados abaixo:

- Utilizar os comitês propostos sobre bases Multibiométricas.
- Realizar um pré-processamento sobre os atributos.
- Propor novos comitês em função de novos classificadores e metaclassificadores.
- Os experimentos foram realizados utilizando 10 cross-validation, dessa forma, é interessante realizar com outras quantidades de iterações.
- Utilizar outras técnicas de validação como: teste de hipóteses.

- Aumentar a quantidade de Algoritmos de mesmo tipo no Comitê.

APÊNDICE A – Comitês que apresentaram os melhores resultados

Como mencionado anteriormente, foram realizados três experimentos para cada comitê, e cada comitê é formado por três algoritmos de AM. Para todos os experimentos foram aplicados os metaclassificadores NB, SMO, IBK, Soma e Voto. As configurações de cada comitê podem ser vistas nas tabelas 14, 15 e 16, que apresentam as configurações para os comitês formados por três, seis e doze algoritmos respectivamente. Os comitês são nomeados de acordo com os algoritmos que o compõe (através de suas abreviaturas) e a quantidade de cada algoritmos (que é representada através do numeral que indica a quantidade de algoritmo daquele tipo no comitê).

Para o algoritmo Árvore de decisão (AD) os parâmetros que tiveram seus valores modificados foram: *ConfidenceFactor* (usado para poda), *Unpruned* (se a poda será executada ou não) e *MinNumObj* (define o número mínimo de instâncias para cada nó terminal da árvore). O algoritmo K-NN teve apenas o parâmetro K modificado, que refere-se ao número de vizinhos mais próximos. No algoritmo MLP foram modificados os parâmetros: *LearningRate* (taxa de aprendizado), *Momentum* (dinâmica aplicada aos pesos durante a atualização que aumenta a velocidade de treinamento da rede neural e reduz o perigo de instabilidade) e *TrainingTime* (número de épocas para o treinamento).

Tabela 14: Configurações dos comitês com três algoritmos de AM que obtiveram os melhores resultados para Taxa de Aprendizado e Desvio Padrão

Comitês	Algoritmo	Parâmetro	Valor
	AD 1	ConfidenceFactor	0.2
		Unpruned	false
		MinNumObj	1
		ConfidenceFactor	0

3 AD	AD 2	Unpruned	true
		MinNumObj	1
	AD 3	ConfidenceFactor	0.2
		Unpruned	false
3 K-NN	KNN 1	MinNumObj	2
		K	1
		K	3
3 MLP	MLP 1	K	7
		LearningRate	0.2
		Momentum	0.5
	MLP 2	TrainingTime	400
		LearningRate	0.25
		Momentum	0.6
	MLP 3	TrainingTime	300
		LearningRate	0.2
		Momentum	0.5
1 MLP, 1K-NN e 1 AD	MLP	TrainingTime	500
		LearningRate	0.3
		Momentum	0.8
	KNN	TrainingTime	600
		K	3
		ConfidenceFactor	0.3
2 AD e 1 K-NN	AD 1	Unpruned	false
		MinNumObj	1
		ConfidenceFactor	0.25
		Unpruned	true
	AD 2	MinNumObj	3
		K	3
	AD 1	ConfidenceFactor	0.3
		Unpruned	false
MinNumObj		2	
ConfidenceFactor		0.25	

2 AD e 1 MLP	AD 2	Unpruned	true
		MinNumObj	3
	MLP	LearningRate	0.3
		Momentum	0.8
		TrainingTime	600
2 K-NN e 1 AD	KNN 1	K	1
	KNN 2	K	3
	AD 1	ConfidenceFactor	0.2
		Unpruned	false
		MinNumObj	1
2 K-NN e 1 MLP	KNN 1	K	3
	KNN 2	K	7
	MLP	LearningRate	0.3
		Momentum	0.8
		TrainingTime	600
2 MLP e 1 AD	MLP 1	LearningRate	0.2
		Momentum	0.5
		TrainingTime	400
	MLP 2	LearningRate	0.2
		Momentum	0.5
		TrainingTime	300
	AD 1	ConfidenceFactor	0.2
Unpruned		false	
2 MLP e 1 K-NN	MLP 1	MinNumObj	0
		LearningRate	0.2
		Momentum	0.5
	MLP 2	TrainingTime	400
		LearningRate	0.2
		Momentum	0.5
	KNN	TrainingTime	300
K		1	

Tabela 15: Configurações dos comitês com seis algoritmos de AM que obtiveram os melhores resultados para Taxa de Aprendizado e Desvio Padrão

Comitês	Algoritmo	Parâmetro	Valor
6 AD	AD 1	ConfidenceFactor	0.5
		Unpruned	false
		MinNumObj	1
	AD 2	ConfidenceFactor	0.25
		Unpruned	true
		MinNumObj	2
	AD 3	ConfidenceFactor	0.35
		Unpruned	false
		MinNumObj	2
	AD 4	ConfidenceFactor	0.35
		Unpruned	false
		MinNumObj	3
	AD 5	ConfidenceFactor	0.95
		Unpruned	false
		MinNumObj	4
	AD 6	ConfidenceFactor	0.25
		Unpruned	true
		MinNumObj	1
6 K-NN	KNN 1	K	1
	KNN 2	K	3
	KNN 3	K	5
	KNN 4	K	7
	KNN 5	K	8
	KNN 6	K	10
	MLP 1	LearningRate	0.2
		Momentum	0.5
		TrainingTime	400
	MLP 2	LearningRate	0.2
		Momentum	0.6
		TrainingTime	400
		LearningRate	0.25

6 MLP	MLP 3	Momentum	0.5	
		TrainingTime	500	
		LearningRate	0.25	
	MLP 4	Momentum	0.6	
		TrainingTime	450	
		LearningRate	0.2	
	MLP 5	Momentum	0.5	
		TrainingTime	500	
		LearningRate	0.15	
	MLP 6	Momentum	0.6	
		TrainingTime	350	
2 MLP, 2 K-NN e 2 AD	MLP 1	LearningRate	0.2	
		Momentum	0.5	
		TrainingTime	400	
	MLP 2	LearningRate	0.2	
		Momentum	0.6	
		TrainingTime	400	
	KNN 1	K	1	
	KNN 2	K	4	
		ConfidenceFactor	0.4	
	AD 1	Unpruned	false	
MinNumObj		0		
ConfidenceFactor		0.25		
AD 2	Unpruned	true		
	MinNumObj	1		
4 AD e 2 K-NN	AD 1	ConfidenceFactor	0.5	
		Unpruned	false	
		MinNumObj	2	
	AD 2	ConfidenceFactor	0.25	
		Unpruned	true	
		MinNumObj	3	
	AD 3	ConfidenceFactor	0.4	
		Unpruned	false	
		MinNumObj	3	
			ConfidenceFactor	0.4

	AD 4	Unpruned	false
		MinNumObj	false
	KNN 1	K	3
	KNN 2	K	7
4 AD e 2 MLP		ConfidenceFactor	0.4
	AD 1	Unpruned	false
		MinNumObj	0
		ConfidenceFactor	0.25
	AD 2	Unpruned	true
		MinNumObj	1
		ConfidenceFactor	0.28
	AD 3	Unpruned	false
		MinNumObj	1
		ConfidenceFactor	0.28
	AD 4	Unpruned	false
		MinNumObj	2
		LearningRate	0.2
	MLP 1	Momentum	0.6
	TrainingTime	400	
	LearningRate	0.2	
MLP 2	Momentum	0.7	
	TrainingTime	400	
4 K-NN e 2 AD	KNN 1	K	1
	KNN 2	K	3
	KNN 3	K	5
	KNN 4	K	7
		ConfidenceFactor	0.5
	AD 1	Unpruned	false
		MinNumObj	1
		ConfidenceFactor	0.25
AD 2	Unpruned	true	
	MinNumObj	2	
	KNN 1	K	1
	KNN 2	K	1
	KNN 3	K	3

4 K-NN e 2 MLP	KNN 4	K	5
	MLP 1	LearningRate	0.2
		Momentum	0.5
		TrainingTime	400
	MLP 2	LearningRate	0.2
		Momentum	0.6
TrainingTime		400	
4 MLP e 2 AD	MLP 1	LearningRate	0.25
		Momentum	0.7
		TrainingTime	500
	MLP 2	LearningRate	0.25
		Momentum	0.8
		TrainingTime	500
	MLP 3	LearningRate	0.3
		Momentum	0.7
		TrainingTime	600
	MLP 4	LearningRate	0.3
		Momentum	0.8
		TrainingTime	450
	AD 1	ConfidenceFactor	0.5
		Unpruned	false
		MinNumObj	1
AD 2	ConfidenceFactor	0.25	
	Unpruned	true	
	MinNumObj	2	
4 MLP e 2 K-NN	MLP 1	LearningRate	0.2
		Momentum	0.6
		TrainingTime	400
	MLP 2	LearningRate	0.2
		Momentum	0.7
		TrainingTime	400
	MLP 3	LearningRate	0.2
		Momentum	0.6
		TrainingTime	500
		LearningRate	0.2

MLP 4	Momentum	0.7
	TrainingTime	350
KNN 1	K	1
KNN 2	K	5

Tabela 16: Configurações dos comitês com doze algoritmos de AM que obtiveram os melhores resultados para Taxa de Aprendizado e Desvio Padrão

Comitês	Algoritmo	Parâmetro	Valor
12 AD	AD 1	ConfidenceFactor	0.4
		Unpruned	false
		MinNumObj	1
	AD 2	ConfidenceFactor	0.25
		Unpruned	true
		MinNumObj	1
	AD 3	ConfidenceFactor	0.4
		Unpruned	false
		MinNumObj	1
	AD 4	ConfidenceFactor	0.28
		Unpruned	false
		MinNumObj	2
	AD 5	ConfidenceFactor	0.36
		Unpruned	false
		MinNumObj	3
	AD 6	ConfidenceFactor	0.4
		Unpruned	false
		MinNumObj	4
	AD 7	ConfidenceFactor	0.25
		Unpruned	true
		MinNumObj	1
	AD 8	ConfidenceFactor	0.25
		Unpruned	true
		MinNumObj	2

	AD 9	ConfidenceFactor	0.25
		Unpruned	true
		MinNumObj	3
	AD 10	ConfidenceFactor	0.25
		Unpruned	true
		MinNumObj	4
	AD 11	ConfidenceFactor	0.5
		Unpruned	false
		MinNumObj	4
	AD 12	ConfidenceFactor	0.4
		Unpruned	false
		MinNumObj	1
12 K-NN	KNN 1	K	2
	KNN 2	K	5
	KNN 3	K	5
	KNN 4	K	9
	KNN 5	K	11
	KNN 6	K	3
	KNN 7	K	3
	KNN 8	K	3
	KNN 9	K	5
	KNN 10	K	7
	KNN 11	K	9
	KNN 12	K	11
	MLP 1	LearningRate	0.2
		Momentum	0.5
		TrainingTime	400
	MLP 2	LearningRate	0.2
		Momentum	0.5
		TrainingTime	300
	MLP 3	LearningRate	0.2
		Momentum	0.6
		TrainingTime	500
	MLP 4	LearningRate	0.2
		Momentum	0.5

12 MLP	MLP 5	TrainingTime	350
		LearningRate	0.2
		Momentum	0.5
	MLP 6	TrainingTime	300
		LearningRate	0.2
		Momentum	0.6
	MLP 7	TrainingTime	400
		LearningRate	0.3
		Momentum	0.5
	MLP 8	TrainingTime	350
		LearningRate	0.3
		Momentum	0.5
MLP 9	TrainingTime	400	
	LearningRate	0.3	
	Momentum	0.6	
MLP 10	TrainingTime	450	
	LearningRate	0.3	
	Momentum	0.5	
MLP 11	TrainingTime	350	
	LearningRate	0.3	
	Momentum	0.5	
MLP 12	TrainingTime	450	
	LearningRate	0.3	
	Momentum	0.6	
MLP 1	TrainingTime	450	
	LearningRate	0.3	
	Momentum	0.8	
MLP 2	TrainingTime	600	
	LearningRate	0.3	
	Momentum	0.85	
MLP 3	TrainingTime	500	
	LearningRate	0.3	
	Momentum	0.9	
MLP 4	TrainingTime	700	
	LearningRate	0.3	
	Momentum	0.9	

4 MLP, 4 AD e 4 K-NN	MLP 4	Momentum	0.8
		TrainingTime	550
		ConfidenceFactor	0.5
	AD 1	Unpruned	false
		MinNumObj	2
		ConfidenceFactor	0.25
	AD 2	Unpruned	true
		MinNumObj	2
		ConfidenceFactor	0.5
	AD 3	Unpruned	false
		MinNumObj	3
		ConfidenceFactor	0.45
	AD 4	Unpruned	false
	MinNumObj	3	
KNN 1	K	5	
KNN 2	K	7	
KNN 3	K	10	
KNN 4	K	1	
8 AD e 4 K-NN		ConfidenceFactor	0.5
	AD 1	Unpruned	false
		MinNumObj	1
		ConfidenceFactor	0.25
	AD 2	Unpruned	true
		MinNumObj	2
		ConfidenceFactor	0.5
	AD 3	Unpruned	false
		MinNumObj	3
		ConfidenceFactor	0.42
	AD 4	Unpruned	false
		MinNumObj	4
		ConfidenceFactor	0.5
AD 5	Unpruned	false	
	MinNumObj	5	
	ConfidenceFactor	0.5	
AD 6	Unpruned	false	

		MinNumObj	6
		ConfidenceFactor	0.25
	AD 7	Unpruned	true
		MinNumObj	3
		ConfidenceFactor	0.25
	AD 8	Unpruned	true
		MinNumObj	4
	KNN 1	K	3
	KNN 2	K	5
	KNN 3	K	7
	KNN 4	K	9
		ConfidenceFactor	0.4
	AD 1	Unpruned	false
		MinNumObj	1
		ConfidenceFactor	0.25
	AD 2	Unpruned	true
		MinNumObj	1
		ConfidenceFactor	0.4
	AD 3	Unpruned	false
		MinNumObj	1
		ConfidenceFactor	0.25
	AD 4	Unpruned	false
		MinNumObj	2
		ConfidenceFactor	0.35
	AD 5	Unpruned	false
		MinNumObj	3
		ConfidenceFactor	0.
	AD 6	Unpruned	false
		MinNumObj	4
		ConfidenceFactor	0.25
	AD 7	Unpruned	true
		MinNumObj	1
		ConfidenceFactor	0.25
	AD 8	Unpruned	true
		MinNumObj	2
8 AD e 4 MLP			

	MLP 1	LearningRate	0.2
		Momentum	0.5
		TrainingTime	400
	MLP 2	LearningRate	0.2
		Momentum	0.55
		TrainingTime	300
	MLP 3	LearningRate	0.2
		Momentum	0.6
		TrainingTime	500
	MLP 4	LearningRate	0.2
		Momentum	0.5
		TrainingTime	350
8 K-NN e 4 AD	KNN 1	K	3
	KNN 2	K	5
	KNN 3	K	7
	KNN 4	K	10
	KNN 5	K	12
	KNN 6	K	3
	KNN 7	K	3
	KNN 8	K	4
		ConfidenceFactor	0.5
	AD 1	Unpruned	false
		MinNumObj	2
		ConfidenceFactor	0.25
	AD 2	Unpruned	true
		MinNumObj	2
		ConfidenceFactor	0.5
	AD 3	Unpruned	false
	MinNumObj	3	
	ConfidenceFactor	0.4	
AD 4	Unpruned	false	
	MinNumObj	4	
	KNN 1	K	1
	KNN 2	K	1
	KNN 3	K	3

8 K-NN e 4 MLP	KNN 4	K	5
	KNN 5	K	9
	KNN 6	K	1
	KNN 7	K	1
	KNN 8	K	1
		LearningRate	0.2
	MLP 1	Momentum	0.5
		TrainingTime	400
		LearningRate	0.2
	MLP 2	Momentum	0.5
		TrainingTime	300
		LearningRate	0.2
	MLP 3	Momentum	0.5
		TrainingTime	500
		LearningRate	0.2
	MLP 4	Momentum	0.5
	TrainingTime	350	
8 MLP e 4 AD		LearningRate	0.25
	MLP 1	Momentum	0.7
		TrainingTime	500
		LearningRate	0.25
	MLP 2	Momentum	0.75
		TrainingTime	400
		LearningRate	0.25
	MLP 3	Momentum	0.8
		TrainingTime	600
		LearningRate	0.25
	MLP 4	Momentum	0.7
		TrainingTime	450
		LearningRate	0.25
	MLP 5	Momentum	0.75
		TrainingTime	400
		LearningRate	0.25
MLP 6	Momentum	0.8	
	TrainingTime	500	

		LearningRate	0.3
	MLP 7	Momentum	0.7
		TrainingTime	450
		LearningRate	0.3
	MLP 8	Momentum	0.75
		TrainingTime	500
		ConfidenceFactor	0.5
	AD 1	Unpruned	false
		ReduceErrorPruning	1
		ConfidenceFactor	0.25
	AD 2	Unpruned	true
		ReduceErrorPruning	1
		ConfidenceFactor	0.5
	AD 3	Unpruned	false
		ReduceErrorPruning	2
		ConfidenceFactor	0.35
	AD 4	Unpruned	false
		ReduceErrorPruning	3
		LearningRate	0.5
	MLP 1	Momentum	0.7
		TrainingTime	500
		LearningRate	0.25
	MLP 2	Momentum	0.75
		TrainingTime	400
		LearningRate	0.25
	MLP 3	Momentum	0.8
		TrainingTime	600
		LearningRate	0.25
	MLP 4	Momentum	0.7
		TrainingTime	450
		LearningRate	0.25
8 MLP e 4 K-NN	MLP 5	Momentum	0.75
		TrainingTime	400
		LearningRate	0.25
	MLP 6	Momentum	0.8

	TrainingTime	500
	LearningRate	0.3
MLP 7	Momentum	0.7
	TrainingTime	450
	LearningRate	0.3
MLP 8	Momentum	0.75
	N	500
KNN 1	K	1
KNN 2	K	3
KNN 3	K	5

Referências

- AHA, D.; KIBLER, D. Instance-based learning algorithms. *Machine Learning*, v. 6, p. 37–66, 1991.
- BITTENCOURT, V. G. *Aplicação de técnicas de aprendizado de máquina no reconhecimento de classes estruturais de proteínas*. Dissertação (Mestrado) — Dissertação de Mestrado, Universidade federal do Rio Grande do Norte., 2005.
- CAVALCANTE, A. L. S.; BACCI, M. D.; HOKAMA, M. L. *Assinatura de Documentos Digitais através da Biometria no Exército Brasileiro*. Dissertação (Mestrado), 2005.
- CHO, S. B. Pattern recognition with neural networks combined by genetic algorithm. In: . [S.l.]: In Fuzzy sets and Systems, 1999.
- COSTA, J. A. F. Classificação automática e análise de dados por redes neurais auto-organizáveis, phd thesis. In: . [S.l.]: Universidade de Campinas, 1999.
- CRAVEN, M.; SHAVLIK, J. Machine learning approaches to gene recognition. In: . [S.l.: s.n.], 1994.
- CUBIST, . R. R. *The University of Waikato*. 1998. Acessado em: Agosto de 2012. Disponível em: <<http://www.rulequest.com/cubist-info.html>>.
- FACELI, K. Inteligencia artificial: Uma abordagem de aprendizado de máquina. In: . [S.l.]: Rio de Janeiro. LCT, 2011.
- FORTES, L. O. Aplicação de técnicas de aprendizado para o controle inteligente de veículos autônomos. In: . [S.l.]: Trabalho de Conclusão de Curso, Universidade do Vale do Rio dos Sinos, 2001.
- FREUND, Y.; SCHAPIRE, S. Experiments with a new boosting algorithm. In: . [S.l.: s.n.], 1996. p. 56–148.
- FUMERA, G.; ROLI, F. *A theoretical and experimental analysis of linear combiners for multiple classifier systems*. Dissertação (Mestrado), 2005.
- FURTADO, V. Tecnologia e gestão da informação na segurança pública. In: . [S.l.]: Rio de Janeiro. Garamond, 2002.
- HAYKIN, S. Redes neurais, princípios e prática. In: . [S.l.]: 2^a edição, 2001.
- HEINEN, M. R.; S., O. F. *Biometria Comportamental: Pesquisa e desenvolvimento de um sistema de autenticação de usuários utilizando assinaturas manuscritas*. 2002. Acessado em: junho de 2012. Disponível em: <<http://www.fortec.edu.br/mkt/artigo5.pdf>>.

KUNCHEVA, L. I. Combining pattern classifiers: Methods and algorithms. In: . [S.l.]: Mitchell, T. (1997). Machine Learning, 2004.

LIU S. E SILVERMAN, M. *A Practical Guide to Biometric Security Technology*, IEEE Computer Society. 2001. Acessado em: Junho de 2012. Disponível em: <www.computer.org/itpro/homepage/JanFeb/security3.htm/>.

LORENA, A. C.; CARVALHO, A. C. P. L. F. *Uma Introdução às Support Vector Machines*. 2007. Acessado em: Junho de 2012. Disponível em: <http://seer.ufrgs.br/index.php/rita/article/viewFile/rita_v14_n2_p43-67/3543/>.

MAGALHÃES, P. S. *Biometria e autenticação*. 2003. Acessado em: julho de 2012. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/2184/1/capsi.pdf>>.

MARTINS, L. D. *Biometria de Digitais*. Dissertação (Mestrado) — UFRJ, 2007.

MCCULLOCK, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. In: . [S.l.: s.n.], 1943.

MITCHELL, t. Machine learning. In: . [S.l.: s.n.], 1997.

NASCIMENTO, D. S. C. *Configuração Heterogênea de Ensembles de Classificadores: Investigação em Bagging, Boosting e MultiBoosting*. Dissertação (Mestrado) — Dissertação de Mestrado, Universidade de Fortaleza - CE, 2009.

OLIVEIRA, V. F. *Biometria Voiceprints*. 2007. Acessado em: Julho de 2012. Disponível em: <http://www.gta.ufrj.br/ensino/eel879/trabalhos_v120072/viviane/PropriedadesAcusticasdoSin>

PATRICK, E. A.; III, F. P. F. *A generalized k-nearest neighbor rule, information and Control*. Dissertação (Mestrado), 1970.

PIMENTA, A. *et al. WEKA-G: Mineração de dados paralela em grades computacionais*. 2009. Acessado em: junho de 2012. Disponível em: <http://www.fsma.edu.br/si/edicao4/FSMA_S120092Estudentil1.pdf>.

POZZER, C. T. Aprendizado por árvores de decisão. In: . [S.l.]: Relatório para disciplina de Programação de Jogos 3D, Universidade Federal de Santa Maria, 2006.

PROENÇA, H. P. Apresentação sobre o algoritmo de am k-nn. In: . [S.l.]: Universidade da Beira Interior, 2008.

QUILAN, R.; KOHAVI, R. *Decision Tree Discovery*. Dissertação (Mestrado), 1970.

ROMANO, S. M. V. *Sistemas Biométricos aplicados a Segurança da Informação: uma abordagem conceitual sobre os principais dados biométricos*. 2002. Acessado em: julho de 2012. Disponível em: <<http://osorio.wait4.org/Signatures/Heinen-Osorio-Infocomp2004.pdf/>>.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. In: . [S.l.: s.n.], 1958.

SEEWALD, A. *Towards a theoretical framework for ensemble classification*. Dissertação (Mestrado), 2003.

- SILVA, D. R. P.; M., S. L. *Segurança da Informação: uma reflexão sobre o componente humano*. 2007. Acessado em: julho de 2012. Disponível em: <<http://www.cienciasecognicao.org>>.
- TAN, A. C.; GILBERT, D.; DEVILLE, Y. Multi-class protein fold classification using a new ensemble machine learning approach. *In Genome Informatics*, v. 14, 2003.
- THIAN, N. *Biometric Authentication System*. Dissertação (Mestrado) — Dissertação de Mestrado, USM, Penang, Malásia, 2001.
- VALADAO, R. B. *Biometria - Reconhecimento de Íris*. 2008. Acessado em: julho de 2012. Disponível em: <<http://www.gta.ufrj.br/grad/081/iris/index.html>>.
- VALE, K. M. O. *Uma análise de métodos de distribuição de atributos em comitês de classificadores*. Dissertação (Mestrado) — Dissertação de Mestrado, Universidade federal do Rio Grande do Norte, 2009.
- WAIKATO, W. T. U. of. *The University of Waikato*. 1993. Acessado em: Junho de 2012. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>.
- WOLPERT, D. Staked generalization. In: . [S.l.]: Neural Networks, 1992.