AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ

AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

Ləman Bəhruz qızı Məmmədova

Xəyalə Misir qızı Musayeva

Aytac Həmid qızı Həmidli

Şəfa Ceyhun qızı Əlizadə

ŞİFRLƏNMİŞ TRAFİKİN KLASSİFİKASİYA ÜÇÜN DEEP LEARNİNG METODLARININ TƏTBİQİ

MÖVZUSUNDA

MAGİSTRLİK DİSSERTASİYASI

İxtisas: 060632 – İnformasiya tenologiyaları və sistemləri mühəndisliyi

İxtisaslaşma: İnformasiya texnologiyaları və telekommunikasiya

Elmi rəhbər: **f.r.e.n. dos. Əhmədova S. R.**

BAKI – 2024

MÜNDƏRİCAT

GİRİŞ

İşin aktuallığı. Şifrələnmiş trafik kibertəhlükəsizlik sahəsində getdikcə daha böyük problemə çevrilir, çünki internetdən istifadə hər gün artmaqdadır. Ənənəvi üsullardan istifadə edərək şifrələnmiş trafikin təhlili çətin və ya qeyri-mümkün ola bilər. Bu kiberhücumçulara öz fəaliyyətlərini gizlətməyə və aşkarlanmadan qaçmağa imkan verir. Şifrələnmiş trafik təsnifatında dərin öyrənmənin istifadəsi bu sahədə yaranan problemləri aradan qaldırmağa köməklik edir. Dərin öyrənmə böyük həcmdə məlumatları təhlil etmək və mürəkkəb əlaqələri müəyyən etmək qabiliyyəti ilə seçilir. Ənənəvi üsullardan fərqli olaraq, dərin öyrənmə alqoritmləri verilənlərdəki nümunələri avtomatik öyrənərək şifrələnmiş trafik təhlilində daha effektiv və dəqiq nəticələr əldə etməyə imkan verir. Bu texnologiyanın istifadəsi kibertəhlükəsizlik mütəxəssislərinə öz şəbəkələrini daha yaxşı qorumağa və potensial təhlükələri daha tez müəyyən etməyə kömək edə bilər. Xüsusilə, dərin öyrənmə alqoritmlərinin istifadəsi şifrlənmiş şəbəkə trafikində gizlənmiş zərərli proqramları və ya hücumları aşkar etməyə kömək edə bilər. Bu təşkilatlara istifadəçilərin məlumatlarını və məlumatlarını daha effektiv şəkildə qorumağa imkan verir.

Tədqiqatın məqsədi və vəzifələri. Bu tədqiqatın əsas məqsədi dərin öyrənmə texnologiyalarından istifadə etməklə şifrələnmiş trafikin təsnifatını tamamlamaq və optimallaşdırmaqdır. Bu məqsədə nail olmaq üçün aşağıdakı vəzifələr müəyyən edilmişdir:

Şifrələnmiş trafik nümunələrinin toplanması və təsnif edilməsi:Bu mərhələdə, müxtəlif mənbələrdən şifrələnmiş trafik məlumatları toplanır, toplanan nümunələr ətraflı şəkildə analiz edilərək müxtəlif kateqoriyalara məsələn, müxtəlif protokollar (HTTPS, SSL/TLS və s.), tətbiq növləri üzrə (Veb trafiki, e-poçt, sosial media, fayl ötürməsi və s), xidmət keyfiyyəti (QoS) üzrə) uyğun olaraq tənif edilir. Bu addım, modellərin təlimi üçün zəruri olan geniş və etibarlı məlumat bazasının yaradılmasını təmin edəcəkdir.

Dərin öyrənmə modellərini tətbiq etməklə şifrələnmiş trafik nümunələrinin təhlili və təsnifatı: Bu mərhələdə əsas məqsəd, şifrələnmiş trafik nümunələrinin müxtəlif xüsusiyyətlərini müəyyən edərək onları dəqiq şəkildə təsnif edə bilən modellərin yaradılmasıdır. Çox qatlı sinir şəbəkələri (CNN), rekurrent sinir şəbəkələri (RNN) və transformer arxitekturaları kimi müxtəlif dərin öyrənmə metodları bu işdə istifadə edilə bilər. Müxtəlif Dərin Öyrənmə Arxitekturalarını Müqayisə Edərək Ən Effektiv Modeli Seçmək: Bu addımda müxtəlif dərin öyrənmə arxitekturaları müqayisə edilir. Hər bir modelin performansı ölçülür və nəticələr təhlil edilir. Məqsəd şifrələnmiş trafikin təsnifatı üçün ən yüksək dəqiqliyi və sürəti təmin edən modeli müəyyən etməkdir. Müqayisə üçün meyarlar, məsələn, geri xatırlatma, dəqiqlik, və F1 skoru kimi müxtəlif üsullardan istifadə olunur. Gələcəkdəki tədqiqat yönümləri və nailiyyətlərin proqnozlaşdırılması: Bu hissədə, şifrələnmiş trafikin klassifikasiyası sahəsində gələcəkdəki tədqiqat yönümləri və potensial nailiyyətlər proqnozlaşdırılır.

Tədqiqatın predmeti və obyekti: Dissertasiyanın obyekti Şifrələnmiş trafikin təsnifatında dərin öyrənmənin tətbiqi. Tədqiqat işinin predmeti isə, Şifrələnmiş trafikdəki məlumatların dərin öyrənmə metodu ilə təsnif edilməsi və bu prosesdə mövcud olan problemlərin həll edilməsi.

Tədqiqat metodları. Dissertasiya işində aşağıdakı tədqiqat metodlarından istifadə edilmişdir: Dərin öyrənmə metodu, təsnifatlaşdırma, analiz, aşkarlama, identifikasiya, görüntü təhlili və s. kimi bir çox sahədə araşdırmalar aparmaq üçün tətbiq edilə bilər. DL metodu ilə tədqiqat çərçivəsində mövcud problemi həll etmək üçün nəzarət edilən maşın öyrənmə metodu ilə üç fərqli alqoritmdən istifadə etmək olar: Qərar ağacı alqoritmi, Təsadüfi meşə alqoritmi; XGBoost (təkmilləşdirmə) alqoritmi.

İnternet trafikinin genişmiqyaslı toplanması, şifrələnmiş trafikin ayrılması və Deep Learning modellərinin tətbiqi ilə əlaqədar tədqiqat metodlarını təqdim edir. Bu metodlar MLP, CNN və RNN kimi fərqli dərin öyrənmə alqoritmləri vasitəsilə şifrələnmiş trafikdəki təhlükələri müəyyənləşdirməyə yönəlib. Alqoritmlərin

performansı Doğruluq, Həssaslıq, Rəqəmsal Həssaslıq və F1 Skoru kimi əsas metrikalarla qiymətləndirilir. Gələcəkdəki inkişaf üçün isə kuantum kompüterlərinin dəstəyi, daha güclü alqoritmlər, sürətli hesablama texnologiyaları və mobil klassifikasiya alqoritmlərinin mövcud imkanları ətrafında inkişaf etmək əhəmiyyətli rol oynayacaq.

Elmi yeniliyin elementləri.

1. Kuantum Kompüterlərin İstifadəsi: Deep learning və kuantum kompüterlərinin birliyi, klassifikasiya alqoritmlərinin effektivliyini artırmaq üçün güclü potensiala malikdir. Kuantum kompüterlər daha kompleks məlumatların analizini və klassifikasiyasını mümkün edir.

2. Daha Güclü Alqoritmlər: Deep learning alqoritmlərinin daha da inkişafı, klassifikasiya alqoritmlərinin effektivliyini artırmaq və daha geniş məlumat setlərində daha yaxşı performans göstərmək üçün yeni imkanlar yaradır.

3. Sürətli Hesablama Texnologiyaları: Klassifikasiya alqoritmlərinin sürətli hesablama texnologiyaları ilə birlikdə istifadəsi, real-vaxt təhlükəsizlik tədqiqatlarında əhəmiyyətli bir rol oynayır.

4. Məlumat Artaşdırılması və Kompleksləşdirilməsi: Gələcəkdəki tədqiqatlar, məlumatın diversifikasiyası və kompleksləşdirilməsi ilə bağlı olacaq, bu da daha geniş məlumat setlərində daha yaxşı klassifikasiya tətbiq etməyə imkan verəcəkdir.

5. Avtomatlaşdırılmış Proseslər: Klassifikasiya proseslərinin avtomatikləşdirilməsi və öz-öyrənməli sistemlərin daha geniş yayılması, təhlükəsizlik təşkilatlarının daha effektiv reaksiya göstərməsinə kömək edir.

6. Ətraflı Anormallıq Aşkar Edilməsi: Klassifikasiya alqoritmlərinin daha çox anormallıqları aşkar etməsi, təhlükəsizlik təşkilatlarına daha çox səlahiyyət verən dəqiqlik və effektivlik səviyyələrində işləməsi ilə bağlı olacaqdır.

7. Mobil Klassifikasiya Alqoritmləri: Mobil cihazlar üçün klassifikasiya alqoritmlərinin inkişafı, mobil təhlükəsizlik tətbiqləri üçün əhəmiyyətli bir rol oynayacaqdır.

Praktiki həll. Şifrələnmiş trafik təhlili üçün dərin öyrənmə üsullarının praktiki həllərindən biri şəbəkə təhlükəsizliyi məhsulları və ya xidmətlərində istifadə olunan proqram və ya sistem ola bilər. Məsələn, bir firewall və ya şəbəkə monitorinq aləti şifrələnmiş trafiki təhlil etmək və potensial təhlükələri aşkar etmək üçün dərin öyrənmə modellərindən istifadə edə bilər.

Kibertəhlükəsizlik əməliyyat qrupları tərəfindən istifadə edilən təhlükənin aşkarlanması və cavablandırılması (TIP) sistemləri. Bu sistemlər daim şəbəkə trafikinə nəzarət edir və dərin öyrənmə modellərindən istifadə edərək anormal fəaliyyəti aşkarlayır. Məsələn, TIP sistemi şəbəkədə qeyri-adi məlumat ötürülməsi və ya təhlükəsizlik siyasətlərini pozan davranış kimi müəyyən nümunələri aşkarlaya və təhlil edə bilər. Başqa bir misal, təhlükəsizlik insidentlərini təhlil etmək və onlara cavab vermək üçün istifadə edilən SIEM (Təhlükəsizlik Məlumatı və Hadisələrin İdarə Edilməsi) sistemidir. SIEM sistemi dərin öyrənmə modellərindən istifadə edərək şifrələnmiş trafiki təhlil edə və şəbəkədəki potensial təhlükələri müəyyən etmək üçün qabaqcıl analitika və ağıllı aşkarlama üsullarından istifadə edə bilər.

Müdafiə üçün təqdim edilən nəticələr. Dərin öyrənmə şifrəli trafik təhlilində ənənəvi metodlarla müqayisədə bir çox üstünlüklər təklif edir. O, yüksək həssaslıq və dəqiqliyi təmin edir, böyük verilənləri emal etmək, davamlı öyrənmək, gizli nümunələri aşkar etmək , optimallaşdırma və avtomatlaşdırma kimi qabiliyyətlərə malikdir Bu üstünlüklər göstərir ki, şifrələnmiş trafiki təhlil etmək üçün dərin öyrənmənin istifadəsi təhlükəsizlik mütəxəssislərinə təhdidləri daha effektiv aşkar etməyə və onlara cavab verməyə imkan verir.

Nəticələrin aprobasiyası. 1-2 may 2024-cü il tarixlərdə Azərbaycanın Ümummilli Lideri Heydər Əliyevin anadan olmasının 101-ci ildönümünə həsr olunmuş “Qabaqcıl texnologiyalar və innovasiyalar” mövzusunda Tələbə və Gənc Tədqiqatçıların IX Respublika Elmi-Texniki Konfrans, Bakı, AzTU, 2024.

Nəşrlər. Dissertasiya işi üzrə 2 konfrans materialları çap edilmişdir.

Fəsil I. Şifrələnmiş Trafik və Təsnifatı

1.1. Şəbəkə Trafiki.

Kompüter şəbəkələri kontekstində "trafik" şəbəkədəki qurğular arasında məlumat paketlərinin hərəkətinə aiddir. Bu məlumatlara veb səhifə sorğuları, fayl köçürmələri, e-poçtlar, axın mediası və rəqəmsal ünsiyyətin hər hansı digər formaları kimi müxtəlif növ məlumatlar daxil ola bilər. "Trafik" termini çox vaxt şəbəkə daxilində məlumat axınının həcmini, sxemlərini və xüsusiyyətlərini təsvir etmək üçün istifadə olunur. Şəbəkə trafikini başa düşmək və idarə etmək səmərəli kommunikasiyanı təmin etmək, şəbəkə performansını optimallaşdırmaq və təhlükəsizliyi qorumaq üçün çox vacibdir.

Məlumatın şəbəkədə səmərəli şəkildə hərəkət etməsini və şəbəkənin effektiv şəkildə idarə olunmasını təmin etmək üçün protokollardan istifadə olunur. Protokol rabitə prosesində iki və ya daha çox cihaz arasında məlumat mübadiləsi üçün qaydalar və qaydalar toplusudur. Bu müddəalar məlumatın nə vaxt, hansı formatda və hansı cihazlar arasında ötürülməsini müəyyən edir. Protokollar məlumatın nizamlı və dəqiq şəkildə çatdırılmasını təmin etmək, təhlükəsizliyi və mümkün olan ən yüksək effektivliyi təmin etmək üçün zəruri tədbirləri nəzərə alır. Bundan əlavə, protokollar cihazlar arasında məlumatın inteqrasiyasını və uyğunluğunu təmin edir. Protokollar həm fiziki, həm də şəbəkə səviyyələrində məlumatların çatdırılmasını təmin etmək üçün istifadə olunan standartlardır. Ethernet protokolları məlumatı cihazlar arasında fiziki kabellər vasitəsilə ötürməyə imkan verir. Bu, məlumatın kompüterlər, marşrutlaşdırıcılar və açarlar kimi cihazlar arasında səmərəli şəkildə hərəkət etməsinə imkan verir [5].

Şəbəkə üzərindən məlumatların ötürülməsində istifadə olunan protokollar aşağıdakılardır:

1. İnternet Protokolu (IP): İnternetin əsas protokolu. IP ünvanları məlumat paketlərini mənbə cihazdan təyinat cihazına yönləndirmək üçün istifadə olunur. İki əsas versiya var: IPv4 və IPv6.

2. Transmissiyaya Nəzarət Protokolu (TCP): TCP IP üzərindən məlumat ötürülməsini təmin etmək üçün istifadə edilən protokoldur. Bu, məlumat axınını parçalara ayıran və hədəf cihaza çatana qədər hər bir parçanın müvəffəqiyyətlə çatdırılmasını təmin edən əlaqə əsaslı bir protokoldur.

3. İstifadəçi Datagram Protokolu (UDP): UDP əlaqəsiz rabitə protokoludur. TCP-dən fərqli olaraq, UDP məlumat ötürülməsi zamanı təhlükəsizlik və ya autentifikasiya təmin etmir. Buna görə də, bəzi proqramlarda (məsələn, audio və ya video axını) sürət və aşağı gecikmə vacibdirsə, UDP-yə üstünlük verilir.

4. Hipermətn ötürmə protokolu (HTTP) və HTTPS: Bunlar veb səhifələri və digər internet resurslarını ötürmək üçün istifadə olunan protokollardır. HTTP mətn əsaslı veb səhifələr üçün istifadə edilsə də, HTTPS (HTTP Secure) şifrələnmiş məlumat ötürülməsini təmin etməklə təhlükəsiz internet rabitəsini təmin edir.

5. Fayl ötürmə protokolu (FTP): Faylların bir kompüterdən digər kompüterə ötürülməsinə imkan verən rabitə protokoludur.

6. Simple Mail Transfer Protocol (SMTP) və Post Office Protocol (POP) / Internet Message Access Protocol (IMAP): Bunlar e-poçt göndərmək və qəbul etmək üçün istifadə olunan protokollardır. SMTP e-poçt serverləri arasında əlaqə qurur, POP və IMAP isə e-poçt müştəriləri və serverləri arasında əlaqəni təmin edir [16].

Bu protokollar İnternet üzərindən ünsiyyət üçün əsas infrastrukturu təşkil edir və müxtəlif tətbiq ssenarilərində istifadə olunur. Əgər mətn şifrələnməyibsə və ya şifrələndikdən sonra şifrəsi açılıbsa, bu məlumatdan hansı protokollardan istifadə olunduğunu müəyyən etmək olar. Yuxarıda göstərilən protokolların hamısı aydın mətn (şifrələnməmiş) məlumatı ötürmək üçün istifadə olunur. Bu protokolları açıq mətn məlumatlarında nəzərdən keçirmək mümkündür [25].

1.2. Şifrələnmə prosesində istifadə olunan alqoritimlər və protokollar.

İnternet üzərində dataların(məlumatların) gizliliyini və müdafiəsini təmin etmək, təhlükəsiz bir şəkildə ötürülməsini təmin etmək üçün şifrələnmədən istifadə olunur.

Bu, məlumatların şifrələmə açarı olmayan hər kəs üçün oxunmaz hala düşəcək şəkildə kodlaşdırılmasını təmin edir və məlumatların təhlükəsiz bir şəkildə ötürülməsini sağlayır. Müştəri (məsələn, veb-brauzer) və server arasında təhlükəsiz əlaqə qurulduqda, onlar arasında mübadilə edilən məlumatlar kriptoqrafik alqoritmlərdən istifadə etməklə şifrələnir. Bu prosesdə, orijinal məlumat oxunmaz formata çevrilir. Qəbul edən tərəfdə, şifrələnmiş məlumatın doğru şifrə açarına malik olmaqla orijinal formasına qaytarılması prosesi baş verir. Yalnız doğru şifrə açarına malik olan tərəf, məlumatın şifrələnməsində istifadə olunan alqoritmlərdən istifadə edərək şifrəni aça bilər və əvvəlcədən göndərilmiş məlumatı oxuya bilər [34].

Riyazi prinsiplərə və funksiyalara əsaslanan kriptoqrafik alqoritmlər, məlumatları şifrələmə və deşifrə etmək üçün istifadə olunur. Bu alqoritmlər, məlumatların şifrələnməsi və şifrə açılması üçün lazımi hesablama əməliyyatlarını təyin edir və bu işləri təhlükəsiz və effektiv şəkildə yerinə yetirir. Kriptoqrafik alqoritmlər iki əsas kateqoriyaya bölünür:

1. Simmetrik Alqoritmlər: Bu alqoritmlər eyni açarı istifadə edərək məlumatları həmişə birbaşa (yəni, göndərən və qəbul edən tərəflər arasında) şifrələmək və deşifrə etmək üçün istifadə olunur.

a) DES (Data Encryption Standard): Ən çox istifadə olunan ilk simmetrik şifrələmə alqoritmlərindən biridir. 56-bitlik açarla işləyir və blok-şifrələmə alqoritmidir.

b) AES (Advanced Encryption Standard): DES-in yerinə mövcud olan ən məşhur şifrələmə alqoritmidir. 128, 192 və 256 bit açar uzunluqları ilə işləyə bilər və daha yüksək təhlükəsizlik səviyyəsi təmin edir.

c) RC4 (Rivest Cipher 4): Stream şifrələmə alqoritmidir və ən çox istifadə olunanlardan biri olmuşdur. Ən çox SSL və TLS protokollarında istifadə olunmuşdur, lakin təhlükəli qabaqcıl attacklar səbəbindən artıq tövsiyə olunmur.

d) Blowfish: Blok-şifrələmə alqoritmidir və ən çox istifadə olunanlardandır. Variyantları, özəlliklə ənənəvi DES-ə üstünlük təmin edir və fərqli blok uzunluqlarında işləyə bilər.

e) Twofish: Bruce Schneier tərəfindən tərtib edilmiş bir simmetrik blok şifrələmə alqoritmidir. AES yarışmacısı olaraq seçilmişdir, lakin standartlaşdırılmamışdır [57].

2. Asimmetrik şifrələmə alqoritmləri isə açar çiftləri (public/private key pairs) adlanan iki açarı istifadə edir. Məlumatlar asimmetrik açarın istifadə ediləcəyi əməliyyatlar üçün şifrələnir və yalnız məlumatın sahibi olan şəxsə məxsus olan özəl açarla deşifrə edilir. Asimmetrik alqoritimlər:

a) RSA (Rivest-Shamir-Adleman): Məlumat şifrələnməsi, imzalama və açar mübadiləsi üçün istifadə olunan ən yaygın asimmetrik alqoritmlərdən biridir. Riyazi əməliyyatların effektiv hesablanmasına əsaslanır.

b) DSA (Digital Signature Algorithm): İmzalama üçün nəzərdə tutulmuş olan bu alqoritmdə, məlumatın əsl olduğunu təsdiqləmək üçün istifadə olunur. Əsasən imzalama əməliyyatları üçün məhsuldar olur.

c) Diffie-Hellman Key Exchange (DH): İki tərəfin açarları mübadilə etməsi və təhlükəsiz bir əlaqə qurmaq üçün istifadə olunan protokoldur. İki tərəfin əsasında gizli bir anlaşma əldə etmək üçün istifadə edilir.

d) Elliptic Curve Cryptography (ECC): Daha yüksək təhlükəsizlik səviyyəsi təmin edə bilən daha sürətli və az yer tutan bir asimmetrik alqoritmdir. RSA-ya nisbətən daha qısa açarlar və işlənmə zamanını azaltmaq üçün daha yaxşı effektivdir. Mobil cihazlar və sərin cihazlar üçün uyğundur.

e) ElGamal: Məlumat şifrələnməsi üçün asimmetrik bir alqoritmdir. Diffie-Hellman protokolunun tətbiqatında istifadə edilir və məktublaşma tətbiqlərində də istifadə olunmuşdur. Özü məlumatın şifrələnməsi üçün effektivdir və özəlliklə açar mübadiləsi tətbiqlərində işlənir [43].

Protokollar məlumatların şifrələnməsi və deşifrə edilməsi üçün kriptoqrafik alqoritmlərdən istifadə edir.

1. HTTPS (Hypertext Transfer Protocol Secure): HTTPS, SSL və TLS protokollarını istifadə edir. SSL əvvəlcədən HTTPS üçün inkişaf etmişdir, lakin sonradan TLS protokoluna dəstək olaraq əvəz olunmuşdur. TLS, SSL-in ən son

versiyası olaraq inkişaf etmişdir və əhəmiyyətli dərəcədə daha təhlükəsiz və müasir bir protokoldur. SSL və TLS, məlumatların təhlükəsiz bir şəkildə ötürülməsini və müştəri və server arasında təhlükəsiz bir əlaqənin qurulmasını təmin edir. Əlaqə qurulduğu zaman, server müştəriyə açıq bir SSL/TLS sertifikatı göndərir. Müştəri bu sertifikatı yoxlayır və əgər etibarlıdırsa, müştəri öz açarını göndərir və əlaqə təhlükəsiz bir şəkildə qurulur. Bu sertifikatlar, məlumatların şifrələnməsinə və serverin kimliyinin təsdiqlənməsinə kömək edir. SSL və TLS, əsasən RSA, DSA, ECC kimi asimmetrik alqoritmləri ilə birlikdə AES, 3DES kimi simmetrik alqoritmləri də daxildir. Asimmetrik alqoritmlər, məlumatın açar mübadiləsi və autentifikasiyası üçün istifadə olunur, simmetrik alqoritmlər isə məlumatın əsas şifrələnməsi üçün istifadə olunur [46].

2. SSH (Secure Shell): SSH, genə simmetrik və asimmetrik alqoritmləri ehtiva edir. SSH-da ən çox istifadə olunan simmetrik alqoritmlər arasında AES, Blowfish və 3DES var. Asimmetrik alqoritmlər arasında RSA, DSA və ECDSA var.

3. SFTP (SSH File Transfer Protocol): Bu protokol, SSH-da istifadə edilən təhlükəsiz fayl transfer protokoludur. Ona görə, SFTP də SSH-da istifadə edilən simmetrik və asimmetrik alqoritmlərdən istifadə edir.

4. FTPS (FTP Secure): FTPS, FTP-nin SSL və TLS ilə təhlükəsizləşdirilmiş variantıdır. Ona görə, SSL və TLS protokolları ilə istifadə edilən alqoritmlər bu protokolda istifadə olunur.

5. SMTPS (Simple Mail Transfer Protocol Secure): E-poçtun təhlükəsiz şəkildə ötürülməsi üçün istifadə olunan protokoldur. Bu protokol SSL və ya TLS ilə təhlükəsizləşdirilmişdir və məlumatların şifrələnməsi və təhlükəsiz bir şəkildə ötürülməsi üçün kriptoqrafik alqoritmlərdən istifadə edir.

6. POP3S (Post Office Protocol version 3 Secure) və IMAPS (Internet Message Access Protocol Secure): E-poçtun məktublarını və məlumatlarını təhlükəsiz bir şəkildə almag üçün istifadə olunan protokollardır. SSL və TLS ilə

təhlükəsizləşdirilmişdir və məlumatların şifrələnməsi üçün kriptoqrafik alqoritmlərdən istifadə edir.

7. IPsec (Internet Protocol Security): İnternet protokolları vasitəsilə məlumatların təhlükəsiz bir şəkildə ötürülməsi üçün istifadə olunan bir protokol suitidir. IPsec, VPN (Virtual Private Network) tətbiqlərində və şəbəkə əlaqələrində təhlükəsizlik təmin etmək üçün kriptoqrafik alqoritmlərdən istifadə edir [52].

Əgər məlumatlar şifrələnibsə, həmin məlumatların içərisindəki protokolları müəyyənləşdirmək çətin olur. Çünki şifrələmə prosesi məlumatı fiziki şifrələmə alqoritmləri ilə qoruyur və müdafiə edir. Klassik protokol tanınma üsulları şifrələnmiş məlumatdan protokolları müəyyən etməkdə çətinlik çəkir, çünki bu üsullar aydın mətn məlumatı üzərində işləyir. Ancaq bunu müəyyən etmənin bir neçə metodu var:

1. Port nömrələri: Şifrələnməmiş məlumat üzərində işləyən protokollar ümumiyyətlə müəyyən port nömrələri üzərində işləyir. Məsələn, HTTP protokolu 80-ci portda, HTTPS isə 443-cü portda işləyir. Buna görə də, məlumat paketlərində hansı port nömrəsinin istifadə olunduğunu yoxlayaraq hansı protokolların istifadə olunduğunu müəyyən etmək olar.

2. Trafik analizi: Şifrələnmiş məlumatların davranış və xüsusiyyətlərini təhlil edərək hansı protokollardan istifadə edildiyini müəyyənləşdirmək mümkündür. Məsələn, verilənlərin paketləri arasında bəzi düzgünliklərin, qaydaların varlığı və ya müəyyən növ əlaqələrinin təyin edilməsi, protokollardan istifadənin müəyyənləşdirilməsinə kömək edir.

3. TLS reklamları: Şifrələnmiş trafikdə TLS (Nəqliyyat Layeri Təhlükəsizliyi) istifadə olunursa, bəzi reklamlar TLS bağlantılarının istifadə edildiyini göstərir. Bu bildirişlər TLS protokollarının mövcudluğunu və istifadə olunan versiyanı göstərir və məlumat üçün hansı protokolların istifadə olunduğunu aydınlaşdırır.

4. Machine Learning (Maşın Öyrənmə): Maşın öyrənmə alqoritmləri, şifrələnmiş məlumatlardan hansı protokolların istifadə edildiyini təxmin etmək üçün öyrənə bilər.

Bu alqoritmlər məlumatın özünəməxsus qaydalarını öyrənərək hansı protokolların hansı məlumatlarla əlaqələndiyini müəyyənləşdirməyə kömək edə bilər.

5. Deep Learning: Dərin Öyrənmə Maşın Öyrənmənin bir alt növüdür məlumatları təhlil etmək və öyrənmək üçün kompleks şəbəkə strukturlarından istifadə edir. Bu alqoritmlər, şifrələnmiş məlumatların özünəməxsus xüsusiyyətlərini öyrənərək hansı protokolların istifadə edildiyini təxmin etməyə kömək edir [61].

1.3. Şifrələnmiş trafikin təsnifatı və yaranan çətinliklər.

Şifrələnmiş trafikin təsnifatı, şifrələnmiş məlumatların qruplaşdırılması və klassifikasiya edilməsini ifadə edir. Bu, məlumatların gizliliyini və təhlükəsizliyini təmin etmək üçün əhəmiyyətli bir təhlükəsizlik tədbiridir. Əsasən aşağıdakı proseslərdən ibarətdir:

1. Məlumat Növlərinə Əsaslı Təsnifatı**:** Məlumat növlərinə əsaslı təsnifat, şifrələnmiş trafikdə yer alan məlumatların növlərinə və cürünə uyğun şəkildə qruplaşdırmaq deməkdir. Bu, məlumatların effektiv şəkildə idarə olunması, monitorinqi və təhlükəsizlik tədbirlərinin tətbiq edilməsi üçün əhəmiyyətli bir yoldur. Məsələn:

a) Elektron poçt məlumatı: İnternet və ya digər şəbəkələr üzərindən elektron məktubların, sənədlərin və digər qeyri-mətn məlumatlarının göndərilməsinə və qəbuluna aiddir. Elektron poçt bütün dünyada məlumat və məlumat mübadiləsinə imkan verən ən məşhur ünsiyyət vasitələrindən biridir.

b) Maliyyə Məlumatı: Banklar, müştərilər və ödəniş xidmətləri tərəfindən göndərilən və qəbul edilən məlumatları ehtiva edir. Buraya kredit kartı məlumatları, bank hesabı məlumatları, ödəniş çıxarışları və s. daxildir. Bu məlumatların məxfiliyi son dərəcə vacibdir.

c) Sağlamlıq Məlumatı: Həkimlər, xəstəxanalar və səhiyyə xidmətləri tərəfindən təqdim edilən və qəbul edilən məlumatları ehtiva edir. Buraya klinik tarix, rəylər və

tədqiqatlar, laboratoriya nəticələri və s. daxildir. Sağlamlıq məlumatlarının məxfiliyi və qorunması da son dərəcə vacibdir.

d) Biznes Məlumatı: Şirkətlər, işçilər, müştərilər və tərəflər arasında göndərilən və qəbul edilən məlumatları ehtiva edir. Buraya işgüzar əlaqələr, müqavilələr, çatdırılma prosesləri, müştəri məlumatları və s. daxildir [70].

2. **Funksiya Əsaslı Təsnifat:** Funksiyaya əsaslı təsnifat, məlumatların istifadə olunduğu funksiyalara görə təsnifatını ifadə edir. Bu təsnifat, məlumatların hansı məqsədlərdə istifadə ediləcəyini və onların hansı funksiyaların təmin etmək üçün əhəmiyyətli olduğunu anlamağa kömək edir. Aşağıda funksiyaya əsaslı təsnifatın əsas mərhələləri yer alır:

a) Ağ İdarəetmə: Bu funksiya, bir ağ infrastrukturunun idarə edilməsi və optimallaşdırılması ilə məşğul olan işlərə məruz qalır. Bu, ağ cihazlarının konfiqurasiyasını idarə etmək, bant genişliyini tənzimləmək, trafik yönləndirməsi, şəbəkə təhlükəsizliyi və monitoringi, və s. daxildir. Ağ idarəçiləri, bu məlumatları işləyərək ağ performansını artırmaq və təhlükəsizliyi təmin etmək üçün tədbirlər görə bilər.

b) İstifadəçi Məlumatları: Bu funksiya, istifadəçilərin məlumatlarını idarə etmək və yoxlamaqla məşğul olur. Bu məlumatlar istifadəçi hesablarının məlumatlarını, profil məlumatlarını, giriş və çıxış aktivlərini, parolları və s. daxildir. İstifadəçi məlumatları, istifadəçi identifikasiyası, məxfilik və istifadəçi təcrübəsinin tənzimlənməsi kimi mühüm prosesləri dəstəkləyir.

c) Operativ Proseslər: Bu funksiya, günlük işləri tənzimləmək və yürütmək üçün lazım olan məlumatları əhatə edir. Bu məlumatlar, iş proseslərinin idarə edilməsi, avtomatlaşdırılması və monitorinqi ilə bağlıdır. Operativ məlumatlar, iş axınını, effektivliyi və keyfiyyəti artırmağa kömək edir [82].

3. Gizlilik Səviyyəsinə Əsaslı Təsnifat: Gizlilik səviyyəsinə əsaslı təsnifat, məlumatların gizliliyinin səviyyəsinin müəyyənləşdirilməsi və onların hansı tədbirlər ilə qorunması gerektiyinin anlaşılmasına kömək edir. Bu təsnifat, məlumatların hər

birinin hansı dərəcədə gizli və hassas olduğunu qiymətləndirir və ona görə müvafiq təhlükəsizlik tədbirləri tətbiq edir. Aşağıda gizlilik səviyyəsinə əsaslı təsnifatın əsas mərhələləri verilmişdir:

a) Hassas Məlumatlar: Hassas məlumatlar, ən yüksək gizlilik səviyyəsinə malik olan və müvafiq qorunma tədbirləri tələb edən məlumatlardır. Bu məlumatlar, şəxsi məlumatlar (kimlik nömrələri, doğum tarixləri, mənzillər), finans məlumatlar (kredit kartı nömrələri, bank hesabları), sağlamlıq məlumatları və s. daxildir. Hassas məlumatlar, ən yüksək səviyyədə şifrələnməli, məhdud istifadəçilər tərəfindən yalnız lazımi olduğu halda əlçatan edilməlidir.

b) Gizli Məlumatlar: Gizli məlumatlar, ortalama bir gizlilik səviyyəsinə malik olan və müvafiq qorunma tədbirləri tələb edən məlumatlardır. Bu məlumatlar, müştəri məlumatları (müştəri adları, əlaqə məlumatları), iş məlumatları (iş prosesləri, müqavilələr), və ya iş təsərrüfatının məxfi məlumatları kimi daxildir. Gizli məlumatlar, ümumiyyətlə, məhdud istifadəçilər tərəfindən yalnız lazımi olduğu halda əlçatan edilməlidir.

c) Qeyri-Hassas Məlumatlar: Qeyri-hassas məlumatlar, ən aşağı gizlilik səviyyəsinə malik olan və genə də qorunma tədbirləri tələb edən məlumatlardır. Bu məlumatlar, ümumi təşkilat məlumatları, ümumi statistik məlumatlar və ya ictimaiyyətə açıq məlumatlar kimi daxildir. Qeyri-hassas məlumatlar, genə də məhdud istifadəçilər tərəfindən yalnız lazımi olduğu halda əlçatan edilməlidir, lakin hassas və gizli məlumatlar kimi qorunmaları ehtiyatlılıqla təmin edilməlidir [18].

Şifrələnmiş trafikdə məlumat şifrələndiyi üçün göndərilən məlumatların nümunələri qarışıq ola bilər. Bu, məlumatın mənasını başa düşməyi və təsnif etməyi çətinləşdirir. Özəlliklə, əgər göndərilən məlumatın hansı növə aid olduğunu və ya hansı protokolla göndərildiyini təyin etmək üçün açıq məlumat yoxdursa, məlumatların təsnifatı və anlaşılması daha da çətinləşir. Bu, məlumatları təhlil edən və müəyyən edən şəbəkə monitorinq alətləri və sistemləri üçün bir çətinlik yaradır.

Şifrələnmiş məlumatları açmaq, normal şərtlərdə istifadəçinin məlumatları qorumaq və gizliliyi təmin etmək üçün etmək istəməyəcəyi bir əməliyyatdır. Lakin, bəzən məlumatları açmaq və anlamaq təhlükəsizlik təşkilatları və ya kibertəhlükəsizlik tədqiqatçıları üçün vacib olabilir. Təhlükəsizlik təşkilatları, mümkün hücumları öncədən təmin etmək və məlumat təhlükəsizliyi açıqlarını tapmaq üçün şifrələnmiş məlumatları analiz edə bilər. Bu, potensial təhlükələri öncədən tanımaq və müdafiə strategiyalarını inkişaf etdirmək üçün əhəmiyyətli ola bilər. Şifrələnmiş trafikdə, təhlükəsizlik məqsədləri ilə məlumatların məxfi olması səbəbi ilə verilənlərin az olması ehtimalı yüksəkdir. Bu, modelin mürəkkəb nümunələri dəqiq şəkildə təmsil etmək və ümumiləşdirməkdə çətinlik çəkməsinə səbəb ola bilər. Fərqli protokolların istifadəsi, şifrələnmiş trafikdə müxtəlif strukturların olmasına səbəb olur. Hər bir protokolün öz formatı və quruluşu olduğundan, şifrələnmiş məlumatların içərisində fərqli görünən strukturlar və nizamlar müşahidə olunur. Bu, məlumatları təsnif etməni çətinləşdirir, çünki hər bir protokolün özünəməxsus xüsusiyyətləri və məlumat formatı var [22].

Şifrələnmiş trafikdə məlumatın çeşidlənməsi və təhlili çoxşaxəli, dinamik və mürəkkəb informasiya ilə işləməyi tələb edir. Bu, yüksək performans və effektivlik tələb edən bir prosesdir. Təsnifat və təhlilin sürətli və effektiv olması, məlumatın prosesə çıxarılması, təhlükəli olmayan məsələlərin müəyyən edilməsi və hərəkət edən qeyri-təhlükəli məsələlərin tanınması mühüm məsələdir.

Bu, təhlükəsizlik tədqiqatçıları və təhlükəsizlik təşkilatları üçün çətin bir işdir. Şifrələnmiş trafikdəki məlumatlar geniş və dinamik ola bildiyi üçün bu məlumatı təhlükəsizlik ekspertləri tərəfindən təhlil və müdafiə etmək tələb olunur [64].

Bu incəlikləri idarə etmək üçün güclü alqoritmlər və texnologiyalar istifadə olunur. Buraya dərin öyrənmə, maşın öyrənməsi, davranış analizi və digər texnologiyalar daxildir. Bu texnologiyalar qarışıq məlumatdakı nümunələri tanımaq, potensial riskləri müəyyən etmək və hərəkət edən təhlükələri tez müəyyən etmək üçün istifadə olunur [39].

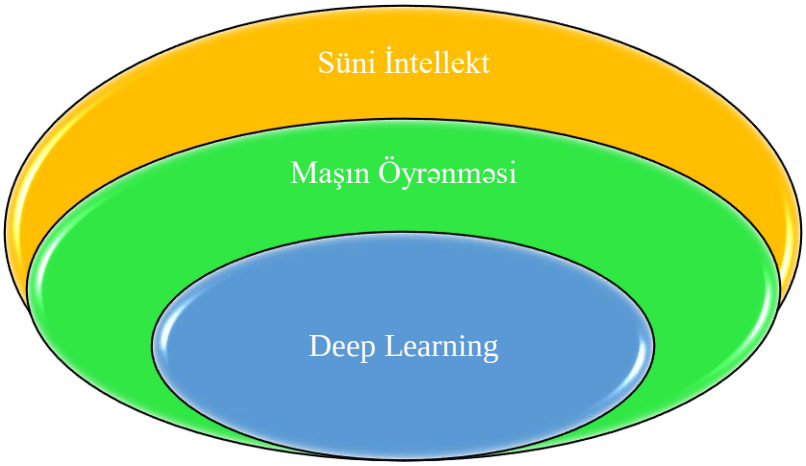
FƏSİL II. Deep Learning metodlarının tədqiqat üsulları.

2.1. Deep Learning (Dərin öyrənmə) metodunun istifadə sahələri.

Deep Learning (Dərin öyrənmə). Dərin öyrənmə maşın öyrənməsinin bir qoludur. Maşın öyrənmənin başlanğıcından bu günə qədər olan dövrdə süni intellektə maraq getdikcə artmış və bu gün ən çox istifadə edilən süni intellekt alqoritmləri olan dərin öyrənmə arxitekturalarının ortaya çıxmasına səbəb olmuşdur. Dərin öyrənmə arxitekturaları ilə birlikdə süni intellekt problemlərini həll etmək üçün bir çox dərin öyrənmə yanaşmaları hazırlanmışdır. Sənaye, tibb, robototexnika, təsvirin işlənməsi, kompüterlə görmə, obyektlərin aşkarlanması, səsin işlənməsi-tanınması, tərcümə, gələcəyi proqnozlaşdırma və maliyyə kimi bir çox sahədə problemlərin ağıllı həlli yollarının hazırlanması nəzərdə tutulur. Deep Learning, maşınların dünyanı anlaması və mürəkkəb problemləri həll etməsinə yönəlmiş süni intellektin inkişafındakı ən populyar yanaşmadır. Dərin öyrənmə metodunda, xarakterizə etmə və çevrilmə aparmaq üçün çoxlu sayda qeyri-xətti emal vahidlərinin təbəqələrindən istifadə edilir. Ardıcıl təbəqələr, əvvəlki təbəqənin çıxışını giriş olaraq qəbul edir və prosesi davam etdirir. DL də, istifadə olunan verilənlər çoxlu sayda xüsusiyyətlərə malikdir. Bunlardan bəziləri yüksək səviyyəli xüsusiyyətlərə, bəziləri isə aşağı səviyyəli xüsusiyyətlərə sahibdir. Aşağı səviyyəli xüsusiyyətlər, yüksək səviyyəli xüsusiyyətlərə malik dərin öyrənmənin xassələrindən törəyir. Dərin öyrənmənin əsasını məhs verilənlərin tam olaraq özündən öyrənmə durur. Deep Learning üsullarına baxdığımız zaman, əllə vasitəsilə aparılan metodlar yerinə verilənləri ən yaxşı şəkildə təmsil edən iyerarxik xüsusiyyətlərin çıxarılması üçün səmərəli alqoritmlər istifadə olunur [3].

Bir sözlə, köhnə alqoritmlər insanlardan asılı formada işlədiyi halda, deep learning alqoritmləri ayırdedici xüsusiyyətləri (özəllikləri) öz-özünə öyrənə bilməsidir. Dərin öyrənmə alqoritmlərinin maşın öyrənmə alqoritmlərindən fərqi; çox böyük məlumatları və bu məlumatları emal edə biləcək yüksək potensiala sahib aparatlara ehtiyac yaranmasıdır. Yüksək potensiallı aparatlar dərin öyrənmənin inkişafı üçün

səmərəli struktura sahibdir və təlim vaxtını qısaldılmasının əsas səbəbi olur. Deep Learning, böyük və mürəkkəb ehtimal modelli sistemlərin öyrədilməsi və bu sistemlərə öyrətmə bacarığı qazandıran çox təsirli bir üsuldur. Ən vacib detal olmasının səbəbi onun verilənlər nümunəsində müxtəlif təbəqələrdən istifadə edilməsidir. Bu fərqli təbəqələrin hamısı ayrı-ayrılıqda əvvəlcədən öyrədilir. Bununla da, ənənəvi yanaşmalardan fərqli olaraq daha uğurlu nəticələr əldə edilir. Dərin öyrənmə tətbiqlərində məlumat dəstinin ölçüsü öyrənmə üçün ən vacib amillərdən biridir. Verilənlər dəsti nə qədər böyük olarsa, öyrənmə bir qədər asan, ancaq təlim müddəti daha uzun olur [6].



Şək. 2.1 Verilənlər dəstin öyrənmə keçidi.

Deep Learning (Dərin öyrənmə) metodunun inkişaf tarixi. Dərin öyrənmə, verilmiş məlumat dəsti ilə nəticələri proqnozlaşdıran çoxsaylı təbəqələrdən ibarət maşın öyrənmə üsuludur. Dərin öyrənmə, maşın öyrənməsi və süni intellekt fərqli mənaları olan terminlərdir. Dərin öyrənmə, maşın öyrənməsi; maşın öyrənməsini süni intellektin bir alt qolu kimi ümumiləşdirmək olar. Süni intellekt insan zəkasına bənzər müxtəlif vəzifələri yerinə yetirən və özünü daim təkmilləşdirən sistemlər və ya maşınlardır. 1950-ci illərdə ortaya çıxan süni intellekt, səhvlərindən dərs ala bilən bir sistem olduğu üçün sistemi daim təkmilləşdirir. Maşın öyrənməsi ilk dəfə 1980-ci illərdən öyrənilməyə başlanılmışdır və verilmiş məlumat dəstini emal etməyi,

proqnozlar verməyi və ya təsnif etməyi əhatə edir. Maşın öyrənmə alqoritmlərində iki növ öyrənmə var: nəzarət edilən və nəzarətsiz öyrənmə.

- Nəzarət olunan öyrənmə etiketlənmiş məlumatlardan öyrənməkdir. Həm giriş, həm də arzu olunan nəticə müəyyən edilir. Proqnozla bağlı dəqiqlik rəyi şəxs tərəfindən verilir.

- Nəzarətsiz öyrənmə etiketsiz müşahidələrdən öyrənmə prosesidir. Nəticə məlumatları ilə təlimə ehtiyac yoxdur. Alqoritmin özü verilənlərdən nəticə çıxarır. Bunu özünüz kəşf etməyiniz gözlənilir [9].

Maşın öyrənməsi, süni intellekt kimi, uzun illər ərzində əhəmiyyətli dərəcədə inkişaf etməmişdir. Onun populyarlığı 1990-cı illərdə data mining ilə artmışdır və dərin öyrənmə 2000-ci illərin əvvəllərində inkişaf etməyə başlamışdır. 90-cı illərdə uğursuzluğun səbəbləri:

- Yetərsiz məlumat dəstləri
- Kompüterlər CPU baxımından çox zəif və qeyri-kafi olması
- Yanlış bir şəkildə ilkləmə (initialization)
- Yanlış qeyri-xətti (aktivasyon) aktivləşdirmə funksiyaları

Zaman keçdikcə texnologiyanın inkişafı kompüterlərin məhsuldarlığını artırmağa imkan vermişdir. Süni intellekt əməliyyatlarında qeyri-kafi olan kompüterlər gücləndirilmiş, verilənlərə çıxış asanlaşdırılmış, alqoritmlərin ehtiyac duyduğu məlumat dəstləri artırılmış, mövcud alqoritmlər təkmilləşdirilmiş və ya yeni alqoritmlər yaradılmışdır. Beləliklə, süni intellekt tətbiqlərində yeni-yeni inkişaflar müşahidə olunmağa başlanılmışdır. Dərin öyrənmə nəzarət altında, yarı nəzarət altında və ya nəzarətsiz həyata keçirilə bilər. Dərin öyrənmədə çoxlu sayda məlumat daxil etməklə fərqli xüsusiyyətləri özündə əhatə edir. Öyrənmə prosesinə nə qədər çox məlumat daxil edilərsə, bir o qədər uğurlu olar. Məlumatlar bir neçə təbəqədən keçir. Üst təbəqələr daha çox təfərrüat çıxaran təbəqələrdir [12].

Deep Learning metodunun növləri və istifadə sahələri. Dərin öyrənmə modellərin üç əsas növü bunlardır:

- Multilayer Perceptron – Çoxlaylı Perseptronlar (Multilayer Perceptrons)
- Konvolyusiya neyron şəbəkələri (Convolutional Neural Networks)
- Təkrarlanan neyron şəbəkələri (Recurrent Neural Networks)

Dərin öyrənmənin istifadə sahələri, üz tanıma sistemlərində, səsin tanınması sistemlərində və avtopilot funksiyası olan və ya sürücüsüz nəqliyyat vasitələrində istifadə olunur. Siqnal sistemlərində kamera qeydlərini daim yoxlamaq əvəzinə, yalnız qeyri-adi hərəkətlər zamanı siqnalizasiya sisteminin işə salınması kimi texnologiyalar dərindən öyrənmə sayəsində mümkün olur. Səhiyyə sektorunda xərçəng araşdırmalarında vaxt itkisini aradan qaldırır. Xərçəngli hüceyrə nümunələrinin təqdim olunduğu dərin öyrənmə alqoritmləri yeni hüceyrələrin xərçəng olub-olmadığını diaqnoz etməkdə daha sürətli və uğurludur. Təsvirin təkmilləşdirilməsində, tövsiyə sistemlərində, bəyəndiyiniz musiqi və film təklifləri, DL üsulları kiber təhdidlərin təhlilində də inkişaf etdirilə bilər. Göstərilənlərdən başqa bir çox misal çəkmək olar [51].

Qonşu ölkəmiz olan Türkiyədə aparılan dərin öyrənmə işlərinə bir nümunə olaraq, ASELSAN-ın Ar-Ge mərkəzində həyata keçirilən silah və təhlükəsizlik sistemləri üçün görüntü və təbii dil işlənməsi sahələrində fəaliyyətlər həyata keçirilir. Digər bir layihə OttOCR, Osmanlı xarakter tanıma sistemidir və Open intelligence layihəsi təsvir və video tanınması üçün dərin öyrənmə API-ləri təklif edir. Süni intellekt proqramlarında ən çox istifadə olunan proqramlaşdırma dilləri Python, c/c++, java və R kimi göstərilə bilər. Dərin öyrənmədə ən çox istifadə olunan dil Python dilidir. Məlumatların vizuallaşdırılması üçün R proqramlaşdırma dilinə üstünlük verilir [15].

Deep Learning metodunun istifadə sahələri. Dərin öyrənmə metodu, təsnifatlaşdırma, analiz, aşkarlama, identifikasiya, görüntü təhlili və s. kimi bir çox sahədə araşdırmalar aparmaq üçün tətbiq edilə bilər. Dərin öyrənmə metodu yarandıqdan sonra görüntü və səs analizində, robot və avtonom nəqliyyat vasitələrində və tibbdə bir çox xəstəliklərin (xərçəng) diaqnozunda istifadə edilməyə başlanılmışdır. Bu metodun tətbiq olunmasının ən böyük səbəbi mürəkkəb problemlərdə belə yüksək

dəqiqlik göstərən nəticələr alına bilməsidir. Bundan əlavə, bəzi audio və video analizlərində insanlardan daha yaxşı nəticələr verə bilir. Hal-hazırda, kameraların sayının artması ilə üz tanınma sistemləri, avtomobil nömrələrinin tanınması sistemləri ilə bağlı araşdırmalar aparılmışdır və uğurlu nəticələr əldə edilmişdir. Bununla yanaşı, Deep Learning sayəsində zərərli şəbəkə trafikinin aşkar edilməsi mümkündür və tələblərə bağlı olaraq uyğun model qurula bilər. Bu gün ortaya çıxan avtonom nəqliyyatlar, Deep Learning vasitəsi ilə avtomobil kameralarından aldığı görüntüləri eyni zamanda təsnif edərək sürücüyə dəstək vahidini təşkil etməkdədir. Bu metodun tətbiq edilməsinin digər bir məqsədi də, qəzaların qarşısı almaqla insan tələfatının azaldılması nəzərdə tutulur. Müxtəlif tətbiq sahələrində istifadə edilməsinə nümunə göstərmək lazım gələrsə, buna misal olaraq istifadəçi ilə danışan sistemlər (Siri), şəkillərin içindəki yazıları mətnə çevirən sistemlər, görüntü keyfiyyəti aşağı olan şəkillərdən istifadə edərək anlaşılan şəkillərin əldə edilməsi və s. göstərmək olar [17].

a. Kompüter Görmə *(Computer Vision).* 2015-ci ildə nəşr olunan bir məqalədə, kompüter görməsi ilə nitq səslərini necə birləşdirilməsi haqqında məlumat verilmişdir. Belə ki, müəyyən bir videoyaya başqa bir nitq söyləməsini oxşar üz və mimika hərəkətləri ilə birləşdirərək təkrarlanan neyron şəbəkələri (RNN) tətbiqi vasitəsilə həyata keçirilmişdir. Bu tətbiqdə Amerika Birləşmiş Ştatlarının keçmiş prezidenti Barak Obamanın video görüntüləri alınıb, daha əvvəllər çıxış etdiyi müxtəlif danışıqları həmin video kadrlara uyğunlaşdırılmışdır [21].

b. Təsnifatlaşdırma *(Classification).* Şəkillərin təsnifatı ilə bağlı çoxlu tədqiqatlar aparılmışdır. Sinifləndirmə üçün istifadə edilən Dərin öyrənmə alqoritmi və üsulları performans dərəcələrində qismən fərqliliklərə səbəb olur. İstifadə olunan müxtəlif üsullar təsnifat performansına qatqı göstərməkdədir. Bunlara bariz nümunə kimi yol nişanlarının, peyk şəkillərinin, üz tanınmasının, hiperspektral məlumatların, üçölçülü təsvirlərin, bir çox xəstəliklərin (hətta xoş və bədxassəli şişlərin), torpaq örtüyü və bitki növlərinin, səs yazılarından sosial siqnalların, Twitter-də duyğuların və s. klassifikasiyasını göstərmək olar [24].

c. Obyektin aşkarlanması *(Object Detection).* Təsvirin təsnifatı və obyekt aşkarlanması oxşar quruluşa malik olduğu görünsə də, bir-birindən fərqlidir. Təsnifatlaşdırmada, mövcud etiketlərə görə görüntülər bir sinifə salınır. Obyekt aşkarlamada isə obyekt təsvirin içərisində axtarılır. Bu obyekt təsvirin harada olduğu təxmin edilir.

d. Səs *(Audi-Wave-Speech).* 2016-cı ildə Google xam audio data yaratmaq üçün WaveNet təqdim etmişdir. Bu model ehtimal və avtoreqressiv quruluşa malik idi. Tətbiqlərdə ingilis və çin dillərində mətni nitqə çevirmək imkanı verir. Bundan əlavə, Baidu tədqiqatçıları tərəfindən səs üçün Deep Speech adlı dərin öyrənmə arxitekturası yaradılmışdır. Deep Speech nitqin tanınması sistemidir. Səs-küylü mühitlərdə belə yaxşı işləmək qabiliyyətinə malikdir [28].

e. Tibbi *(Medical).* Dərin öyrənmə tədqiqatları ilə insan sağlamlığına birbaşa təsir edəcək bir çox yanaşma vasitəsiləri irəli sürülmüşdür. İnsan sağlamlığı üçün əhəmiyyətli olan bir çox məlumatlar, klassifikası, aşkarlama, təsvirin seqmentasiyası və görüntü istehsalı kimi bir çox proseslərdə istifadə olunur. Bu proseslər daha çox sinir sistemləri, ağciyərlər, gözlər, patoloji görüntülər, hüceyrələr, döş qəfəsi, ürək, qarın və əzələ sistemlərində aparılmışdır. Araşdırma apardığımız ədəbiyyatların birində, dərin qıvrımlı sinir şəbəkələrindən istifadə edərək radioloji şəkillərdən diz osteoartritinin dərəcəsini avtomatik ölçmək üçün bir üsul öyrənildiyini verilmişdir. Bu proqram üçün AlexNet və VGG-16 DL alqoritmlərindən istifadə edilmişdir. Performans dərəcəsini artırmaq üçün öz təklif etdikləri xüsusiyyət çıxarma metodundan istifadə edilmişdir [30].

f. Sənaye *(Industrial).* Sənaye sahəsində istifadə ediləcək dərin öyrənmə arxitektura yanaşmaları istehsal, istifadə, istifadə edən və erqonomika baxımından olduqca çox əhəmiyyətlidir. Sənayenin bütün sahələrində istifadə olunacaq kompüter dəstəkli yanaşmalar məhsuldarlığı artırmaqdadır. Sənaye prosesinə nəzarətdə; əsas dəyişənlər, məhsulun keyfiyyətləri, texniki və iqtisadi məhdudiyyətlər çox vacib hesab olunur. Shang və onun həmkarları bu məlumatların effektiv həlli kimi verilənlərə

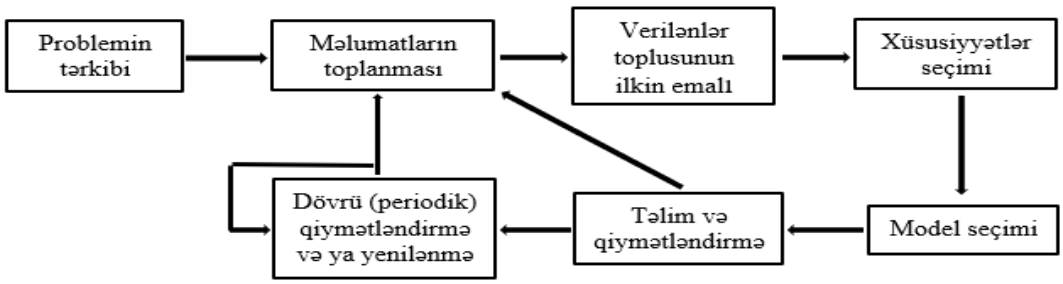
əsaslanan yumşaq sensorlar vasitəsilə bu məlumatların etibarlı və sabit onlayn proqnozlarını yaratmaq üçün bir yanaşma təqdim etmişdir. Bu araşdırmada yumşaq sensorlar yaradaraq ağır dizelin kəsilmə nöqtəsini proqnozlaşdırmaq üçün dərin öyrənmə yanaşması üzərində çalışmalar aparılmışdır [33].

2.2. Deep Learning metodu vasitəsilə problemlərin müəyyən edilməsi.

Trafik təsnifatı iki onillik ərzində ISP-lərdə QoS təminatı və ödənişdən tutmuş təhlükəsizlik divarları və müdaxilənin aşkarlanması sistemlərində təhlükəsizliklə bağlı tətbiqlərə qədər geniş çeşiddə tətbiqlərə tədqiq edilmişdir. Port əsaslı, verilənlər paketinin yoxlanılması və klassik maşın öyrənmə üsulları keçmişdə geniş şəkildə istifadə edilmişdir. Lakin İnternet trafikindəki kəskin dəyişikliklər, xüsusən də şifrələnmiş trafikin artması səbəbindən onların dəqiqliyi azalmışdır. Dərin öyrənmə üsullarının yayılması ilə tədqiqatçılar bu yaxınlarda trafik təsnifatı tapşırığı üçün bu üsulları araşdırmış və yüksək dəqiqlik olduğu haqqında məlumat bildirmişdirlər. Bu tədqiqat işində, biz dərin öyrənmə əsaslı trafik təsnifatı üçün ümumi çərçivə təqdim edirik. Biz tez-tez istifadə olunan dərin öyrənmə üsullarını və onların trafik təsnifatı tapşırıqlarında tətbiqini təqdim edirik. Sonra açıq problemləri, habelə mobil hərəkətinin təsnifatı imkanlarını müzakirə edirik [26].

Şəkil 2.1-də yeddi addımdan ibarət yol hərəkətinin təsnifatı üçün ümumi çərçivəni təsvir edir. Mövcud işlərin əksəriyyəti

çərçivənin hamısını və ya bir hissəsini qəbul edir [37].



Şək. 2.2 Şifrələnmiş trafikin şəbəkə təsnifatının qurulması üçün ümumi sxem.

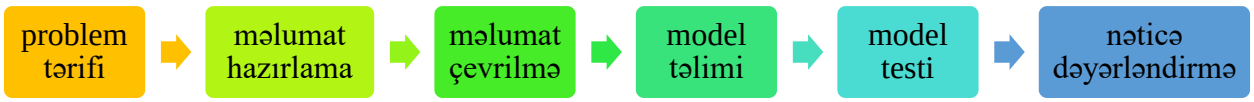
Trafik təsnifatı çox fərqli ssenarilərə tətbiq edilsə də, əksər tədqiqatlar iki geniş yayılmış aspekti bölüşür:

a) təsnifat üçün giriş məlumatları xam paket datası, onun bir hissəsi və ya birbaşa ondan əldə edilən məlumatdır,

b) oxşar ML alqoritmləri istifadə olunur. Burada, diqqəti şifrələnmiş proqram/trafik tipli təsnifatdır.

Maşın öyrənməsindən istifadə etmək üçün əvvəlcə problem müəyyən edilməli və formalaşdırılmalıdır. Bu mərhələdən sonra məlumatların toplanması və toplanmış məlumatların təhlili aparılır. Məlumatlar maşınların anlaya biləcəyi formata çevrilir. Daha sonra məlumatlar təlim və sınaqlara bölünür. Model təlim məlumatları ilə öyrədildikdən sonra test məlumatlarından istifadə edərək modelin performansı qiymətləndirilir. Nəhayət, model statistik üsullarla təhlil edilir, mühüm xüsusiyyətlər və parametrlər müəyyən edilərək model yenidən təklif edilmişdir.

Şəkil 2.2-də maşın öyrənməsi üçün zəruri addımları göstərir [40].



Şək. 2.3 Maşın öyrənmə addımları.

Ədəbiyyatda maşın öyrənmə tətbiqləri üçün müxtəlif növ üsullardan istifadə olunur. Nəzarətli və nəzarətsiz öyrənmə daha çox istifadə olunsa da, yarı nəzarətli və gücləndirici təlim metodlarından da istifadə olunur. Metod və ya üsullar problemə uyğun olaraq müəyyən edilir. Nəzarət olunan təlim metodu üçün istifadə olunan məlumatlar etiketlənməli və təsnif edilməlidir. Bu öyrənmə üsuluna ən çox təsnifat və proqnozlaşdırma məsələlərində üstünlük verilir. Bu üsulda daha çox kNN, Decision Tree, Naive Bayes, SVM, Random Forest alqoritmlərindən istifadə edilir. Bu metodun

ən böyük problemi verilənlərin etiketlənməsidir. Bununla belə, eyni metodologiyadan kiçik dəyişikliklərlə digər təsnifat problemləri üçün də istifadə oluna bilər [14].

Zaman seriyası və statistik xüsusiyyətlər şifrələnməmiş trafik və eyni trafikin şifrəli versiyası üçün bir qədər fərqli ola bilsə də, onlar şifrələmədən asılı olmayaraq da mövcud ola bilər. Beləliklə, şifrələnməmiş trafik üçün bu xüsusiyyətlərdən asılı olan üsullar şifrələnmiş trafiklə də işləyə bilər. Digər tərəfdən, faydalı yük məlumatları və bəzi başlıq məlumatları, məsələn, IPsec tərəfindən şifrələnmiş trafikin dördüncü səviyyə məlumatı şifrələnmiş trafik üçün düz mətndə mövcud olmaya bilər [42].

2.3. Deep Learning metodlarının tədqiqat üsulları vasitəsilə problemlərin təhlili.

Enc xüsusiyyəti yalnız şifrələnmiş trafikə aid olsa da, o, müəyyən şifrələnmiş protokollarla məhdudlaşmır və şifrələnmiş trafikin özünəməxsus xüsusiyyətlərini daha yaxşı təmsil edə bilər. Bu, onu protokola xas xüsusiyyətdən daha çox şifrələnmiş trafik təhlili üçün daha uyğun gəlir. Bir çox müəlliflər həmçinin tələb olunan xüsusiyyətləri çıxarmaq üçün yeni xüsusiyyət yaratma yanaşmalarını və qruplaşdırma metodunu təklif etmişdirlər. Meghdouri və başqa tədqiqatçılar TLS və IPSec protokolları altında şifrələnmiş trafikin yeni təbəqələrarası xüsusiyyət təmsilçiliyini təklif etmişdir. Müəlliflər çıxarış axınının üç rejimini müəyyən etdilər:

- Tətbiq axınları,
- Söhbət axınları və
- Son nöqtə axınları.

Zhang və başqaları 2021-ci ildə TLS/SSL protokolları üçün protokola xas trafik xüsusiyyətləri üçün yeni kodlaşdırma metodu təklif etmişdir. Kodlaşdırma metodu çıxarılan xüsusiyyətləri görüntüyə bənzər məlumat formatına çevirir. Meghdouri və başqalarında yaratdıqları xüsusiyyətlərə baxmayaraq belə bir görüntüyə bənzər xüsusiyyətlər dəsti CNN modelləri ilə qidalanır. Şifrələnmiş trafik üçün özəl

xüsusiyyətlərdir və yalnız TLS/SSL/IPSec protokolları üçün nəzərə alınır. Trafik digər şifrələnmiş protokolları ehtiva etdikdən sonra bu funksiyalar artıq tətbiq olunmur [48].

Aceto və başqa tədqiqatçılar trafik xüsusiyyətləri yaratmaq üçün iki yol təklif etmişdir. Birinci yol sessiyanın 784 faydalı yük baytıdır. İkinci yol, ilk 32 paketin paket istiqamətini, ölçüsünü, TCP pəncərəsinin ölçüsünü və çatma vaxtını hesablamaqdan ibarətdir. Bader və başqaları 2022-ci ildə pro, ilk 32 trafik paketini 5 qrupa (iki istiqamətli paketlər, mənbə, təyinat, əl sıxma paketləri və məlumat ötürmə paketləri) təşkil edən və sonra bu müxtəlif qrup səviyyələrinə əsaslanan yeni xüsusiyyət yaratma yanaşmasını irəli sürmüşdür. Buna uyğun olaraq model təlimi üçün statistik xüsusiyyətlər çıxarılır. Bənzər bir xüsusiyyət qrupunu Bekerman digər müəlliflər protokol-aqnostik xüsusiyyətləri çıxarmaq üçün trafik məlumatlarını 4 qrupa (söhbət pəncərəsi, axın, sessiya, əməliyyat) ayırırlar [2].

Lopez Martin və başqaları 2017-ci ildə hər sessiya axınında trafik paketlərinin uzunluğunun təsirini öyrənmişdir. Onlar hər sessiya axınında trafik paketlərinin 5 ilə 15 paket arasında saxlanması hesablama vaxtı və aşkarlama performansı arasında mübadilə olduğunu təklif etmişdirlər. Hər bir trafik seansında trafik paketlərinin sayının ardıcıl olmasını təmin etdikdən sonra protokol-aqnostik ədədi xüsusiyyət çıxarılması həyata keçirilir. Yaradılmış xüsusiyyətlərin hamısı protokola xas olan xüsusiyyətlərə nisbətən daha sürətli və çıxarılması asan olan protokol-aqnostik ədədi xüsusiyyətlərə qədər uzundur. Bununla belə, yuxarıda müzakirə edildiyi kimi, ənənəvi protokol-aqnostik xüsusiyyətlər yalnız sessiya səviyyəsinin şifrələmə xüsusiyyətlərini nəzərə alınır və paket səviyyəsinin şifrələmə xüsusiyyətlərini nəzərə alınmır. Təklif etdiyimiz Enc xüsusiyyəti ilə digər ən müasir funksiyaların yaradılması və mühəndislik yanaşmaları arasında müqayisə də ümumiləşdirilmişdir [55].

2.4. Deep Learning metodu ilə sistem qurulması.

Şəbəkə trafikinin təsnifatı üçün dərin öyrənmə modellərinin seçilməsinə bir neçə amil təsir göstərir. Ən vacibi isə xüsusiyyətlərin seçimidir. Xüsusiyyətlər hesablama

mürəkkəbliyinə və təsnifat üçün paketlərin sayına (yaddaş mürəkkəbliyi) təsir edən giriş strukturuna və ölçüsünə birbaşa təsir göstərir. Sonra, seçilmiş xüsusiyyətə əsasən uyğun bir model seçilməlidir. Burada yalnız başlıq xüsusiyyətləri əhatə edilmir və yalnız digər xüsusiyyətlərlə birlikdə əhatə edilir. Çünki tək başlıq xüsusiyyətləri təsnifat üçün həmişə kifayət qədər təsirli olmaya bilər.

Başlıqlarda yalnız port nömrəsi, pəncərə ölçüsü və bəzi nadir hallarda xidmət növü (ToS) və ya parçalanma ilə əlaqəli sahələr təsnifat üçün faydalı məlumat verilir. Daxiletmə xüsusiyyətinin seçimi və maşın öyrənmə metodu yüksək dərəcədə əlaqələndirilir. Bundan əlavə, verilənlər bazasının ölçüsü də model seçiminə təsir göstərir [58].

Məsələn, verilənlər bazası kiçik olduqda dərin öyrənmə metodları uyğun gəlmir. Verilənlər dəstinin böyük olduğunu fərz etsək, üç tez-tez istifadə olunan giriş funksiyası və müvafiq uyğun modellər aşağıda təsvir edilmişdir:

Zaman Seriyası+Başlıq: Zaman seriyası xüsusiyyətləri şifrələmədən çətinliklə təsirləndiyi üçün müxtəlif proqramlar və verilənlər dəstləri üçün geniş şəkildə tətbiq edilmişdir. 10 ilə 30 paket arasında olan ilk bir neçə paketin bir çox verilənlər bazasında təsnifat üçün kifayət olduğu bildirilir. Bütün axından nümunə götürülmüş paketlər də ümidverici dəqiqliyə nail olmaq üçün göstərilmişdir.

Klassik ML alqoritmləri və MLP modelləri, giriş ölçüsünü təmsil edən paketlərin sayı az olduqda yaxşı işləyir. Daha çox sayda paket üçün CNN və LSTM-nin daha dəqiq olduğu bildirilir. Kiçik sayda paketlər üçün belə, CNN modeli istifadə olunur. Hesablama mürəkkəbliyi və dərin modellərin təlim müddəti klassik maşın öyrənmə alqoritmlərindən daha yüksəkdir [60].

Yük+Başlıq: Cari şifrələnmiş trafiklərdə əl sıxma məlumatını ehtiva edən ilk bir neçə paket adətən şifrələnmir və təsnifat üçün uğurla istifadə olunur. Girişin yüksək ölçülü olmasına görə (faydalı yükdə çoxlu sayda bayt) klassik ML metodları və MLP yaxşı işləmir.

Belə hallarda, CNN və ya CNN və LSTM birləşməsinin yüksək dəqiqliyə malik olduğu bildirilir. Dəqiqliyi bir qədər yaxşılaşdırmaq üçün yük məlumatı ilə yanaşı zaman sıralarının xüsusiyyətlərindən də istifadə etmək mümkündür, lakin bu, giriş ölçüsünü və ya model seçimini demək olar ki, dəyişdirmir [65].

Statistik Xüsusiyyətlər: Statistik xüsusiyyətlərin sayı və nəticədə daxiletmə ölçüsü məhduddur. Beləliklə, əksər sənədlər bu xüsusiyyətlərə görə klassik ML metodlarından və ya nadir hallarda MLP-dən istifadə edilir. Əksər tədqiqatlar bütün axını müşahidə etməklə statistik xüsusiyyətlər əldə edilsə də, məlumat dəstləri və statistik xüsusiyyətlərin seçimindən asılı olaraq ilk 10 paketdən 180 paketə qədər statistik xüsusiyyətlərin əldə edilməsinin təsnifat üçün kifayət ola biləcəyi göstərilmişdir.

Statistik xüsusiyyətlər bizə klassik ML alqoritmlərinə əsaslanan daha sadə təsnifat qurmağa imkan versələr də, onlayn və sürətli təsnifat üçün uyğun olmaya bilər. Çünki axından etibarlı statistik xüsusiyyətlər əldə etmək üçün kifayət qədər paket tutmalıdır [81].

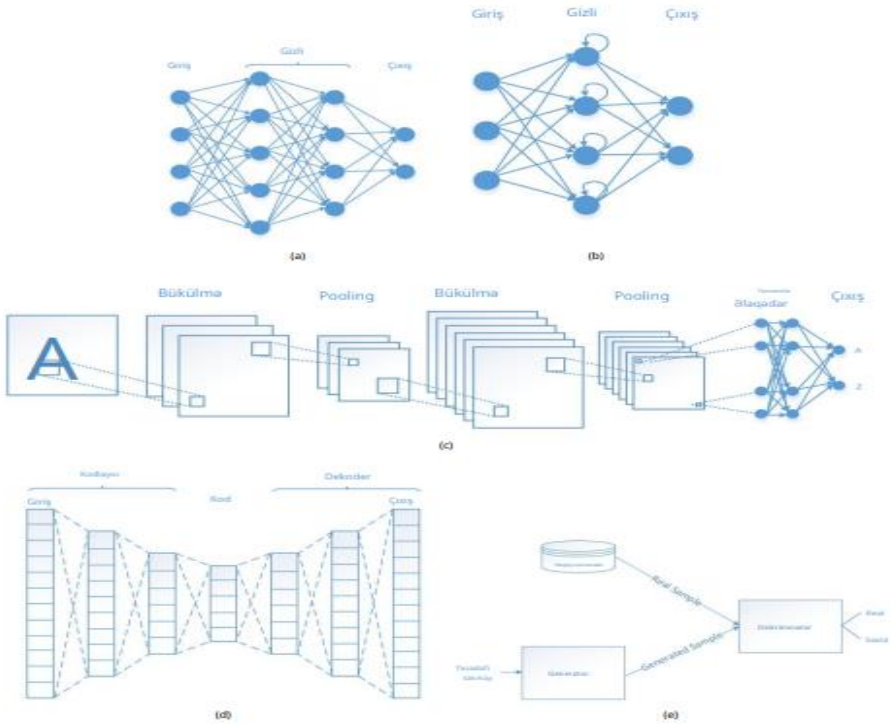
Cədvəldə, müvafiq modelləri və onların xüsusiyyətləri ümumiləşdirilmişdir. Bütün bu yanaşmaların müəyyən bir verilənlər bazası üçün işlədildiyinə heç bir zəmanət yoxdur. Verilənlər kifayət qədər deyilsə, məlumat toplama mərhələsi, seçilmiş xüsusiyyətlər, model təmsil olunmursa, xüsusiyyət, model seçimi mərhələsinə keçmək lazım gələ bilər. Bu yanaşmalar yalnız müəyyən trafiklər üzrə öyrənilmişdir.

QUIC və TLS 1.3 kimi daim artan və gələcək protokolların hərtərəfli tədqiqi hələ aparılmamışdır [66].

Cədvəl 2.1

Model və xüsusiyyətlərin seçilməsi üçün təlimat

Xüsusiyyət	Zaman seriyası+başlıq	Yük+başlıq	Statistik
Model	Klassik ML/MLP/CNN/LSTM	CNN/CNN+LSTM	Klassik ML/MLP
Hesablama mürəkkəbliyi	Aşağı/orta	Yüksək	Aşağı
Lazım olan paketlərin sayı	Orta	Aşağı	Yüksək

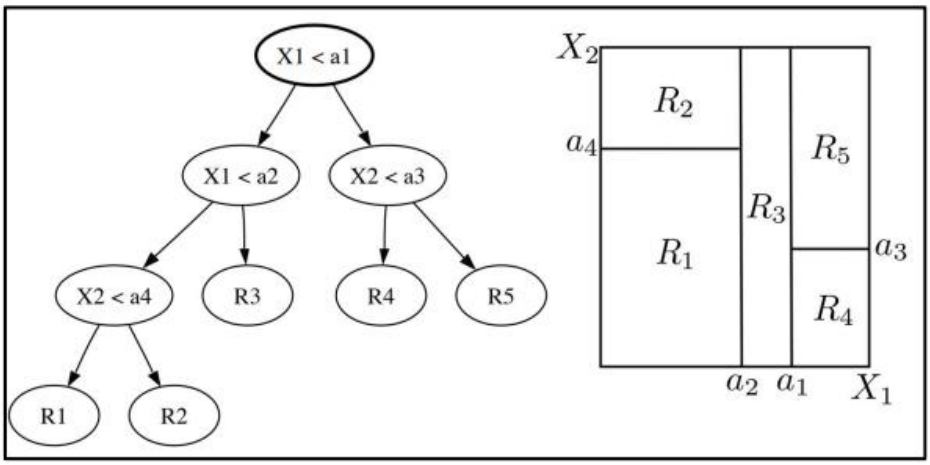


Şək. 2.4 Ümumi dərin öyrənmə modelləri:
a) MLP, b) CNN, c) RNN, d) AE, e) GAN.

DL metodu ilə tədqiqat çərçivəsində mövcud problemi həll etmək üçün nəzarət edilən maşın öyrənmə metodu ilə üç fərqli alqoritmdən istifadə etmək olar.

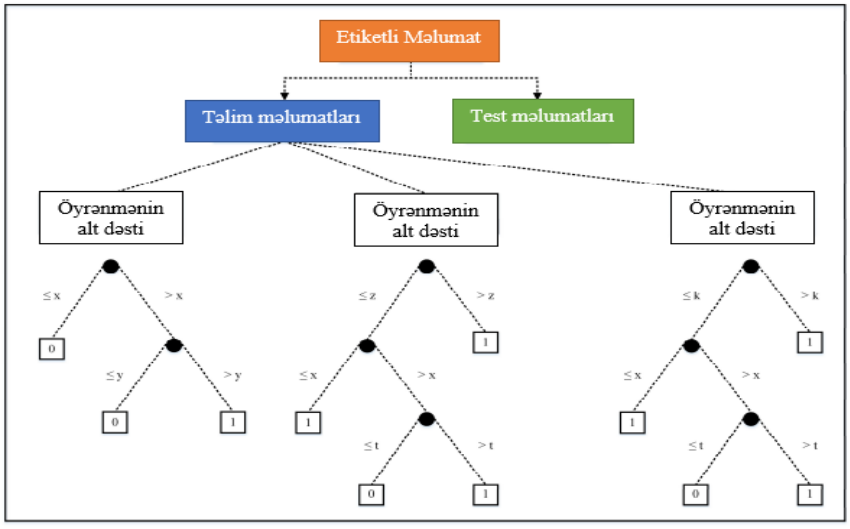
1. Qərar ağacı alqoritmi (Decision tree algorithm);
2. Təsadüfi meşə alqoritmi (Random forest algorithm);
3. XGBoost (təkmilləşdirmə) alqoritmi (XGBoost algorithm) [69].

Qərar ağacı alqoritmi (Decision tree algorithm). Qərar ağacı alqoritmi Quinlan tərəfindən hazırlanmış geniş istifadə olunan maşın öyrənmə alqoritmidir. Qərar ağacı alqoritmi təsnifat, qruplaşma və reqressiya məqsədləri üçün istifadə edilə bilər. Bu üsul vasitəsilə problemin həllində çoxlu qovşaqlar, bunlar arasında əlaqələr və sonda isə qərar ağacı yaradılır. Qərar ağaclarını öyrətmək, qiymətləndirmək tez və şərh etmək isə asandır. Desicion Tree alqoritmi kateqoriyalı və ya ədədi verilənlər üzərində işləmək qabiliyyətinə malikdir [72].



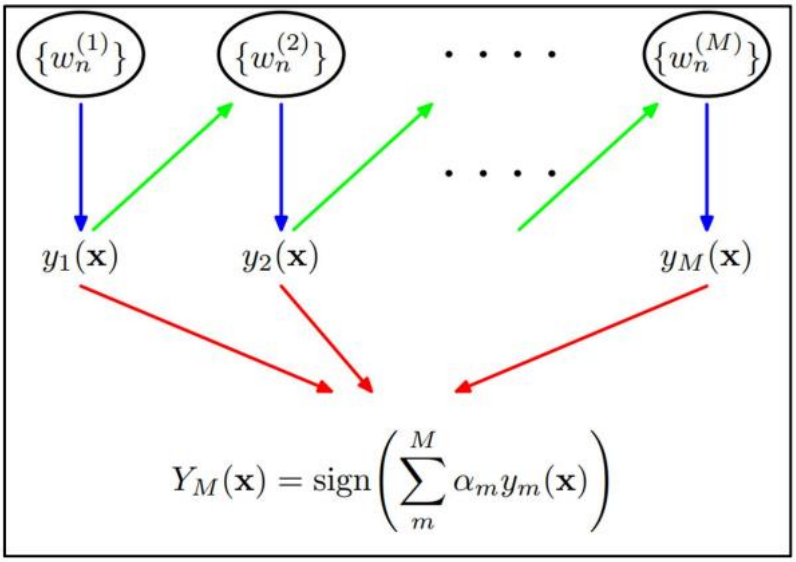
Şək. 2.5 İki ölçülü qərar ağacı (Two-dimensional decision tree).

Təsadüfi meşə alqoritmi (Random forest algorithm). Ansambl metodu daha dəqiq model yaratmaq üçün bir neçə qiymətləndiricinin birləşdirilməsi üsuludur. Random Forest alqoritmi 2001-ci ildə L.Breiman tərəfindən hazırlanmışdır və ansambl üsuludur. Bu üsulda, "bir ağac yaxşıdırsa, çoxlu ağaclar da yaxşıdır" məntiqi ilə qurulur. Bu alqoritmdə qərar ağaclarından istifadə edərək təsadüfi meşə yaradılır. Hər ağac fərqli təlim dəstləri ilə qurulduğu üçün onların strukturları bir-birindən fərqlidir və qərar ağaclarının edə biləcəyi seçimlər məhduddur. Hər bir qərar ağacı bir sinif yaradır. Bütün qərar ağaclarından alınan məlumatlar qiymətləndirilir və son nəticəyə gəlinir. Bu xüsusiyyət sayəsində geniş miqyaslı problemlərdə istifadə edilə bilər [75].



Şək. 2.6 Random Forest alqoritminin strukturu (Random forest structure).

XGBoost (təkmilləşdirmə) alqoritmi (XGBoost algorithm). Boosting, vahid model yaratmaq üçün bir çox öyrənmə modellərini birlikdə istifadə edən kollektiv öyrənmə modelidir. Bu üsul zəif alqoritmləri yenilənmək və ardıcıl öyrətməklə gücləndirir. Gradient Boosting, gücləndirməyə əsaslanan və qərar ağacları kimi zəif proqnozlaşdırma modellərinin ardıcıl partiyasını yaradaraq və öyrətməklə güclü model istehsal edən bir üsuldur. 2001-ci ildə Fridman tərəfindən yaradılmışdır. Hər bir ağac itki funksiyasını azaltmaq üçün əlavə edilir və ardıcıl olaraq əlavə olunan ağaclar itki funksiyasını azaldır. XGBoost Gradient Boosting alqoritminə əsaslanır. Lakin optimallaşdırma ilə performans yaxşılaşmasına nail olunmuşdur. XGBoost açıq mənbəli alqoritmdir və digər maşın öyrənmə alqoritmlərindən on qat daha sürətli işləyə bilər. Paralel və paylanmış hesablama idarəetməsi ilə öyrənmə prosesini çox daha sürətli tamamlayır [78].



Şək. 2.7 Təlim ardıcıllığını təkmilləşdirmə (Upgrade sequence training).

FƏSİL III. Şifrələnmiş trafikin klassifikasiya üçün Deep Learning metodları

3.1. Deep Learning istifadə edərək Mobil Şifrələnmiş Trafik Təsnifatı.

Əl qurğularının kütləvi şəkildə tətbiqi ev və müəssisə şəbəkələri, eləcə də İnternetdən keçən mobil trafik həcminin partlamasına səbəb olur. Trafik Təsnifatı (TC) kimi tanınan bu cür trafik yaradan (mobil) tətbiqlərin nəticə çıxarma prosedurları yüksək qiymətli profil məlumatı üçün imkan yaradır və şübhəsiz ki, mühüm məxfilik problemlərini də aradan qaldırır. Bununla belə, dəqiq təsnifatçıların dizaynı şifrələnmiş protokolların (məsələn, TLS kimi) artan qəbulu ilə daha da pisləşir və paketlərin dərin təftişi kimi yüksək dəqiqlikli yanaşmaların tətbiqinə mane olur. Bundan əlavə, (gündəlik) genişlənən proqramlar dəsti və mobil trafikin hərəkətli-hədəf təbiəti əl ilə və ekspertlər tərəfindən yaradılan xüsusiyyətlərə əsaslanan adi maşın öyrənməsi ilə dizayn həllərini köhnəlmiş olduğunu ehtiva edir [1].

Bu səbəblərə görə, mürəkkəb mobil trafik nümunələrini əks etdirən avtomatik olaraq çıxarılan xüsusiyyətlərə əsaslanan trafik təsnifatlandırıcılarının dizaynı üçün əlverişli strategiya kimi Deep Learning (DL) təklif edilmişdir. Bu məqsədlə, TC-dən fərqli müasir DL texnikaları burada çoxaldılır, bölünür və müqayisə üçün sistematik çərçivəyə, o cümlədən performansın qiymətləndirilməsi iş masasına daxil edilir.

Real insan istifadəçilərinin fəaliyyətinin üç verilənlər bazasına əsaslanaraq, bu DL təsnifatçılarının performansı kritik şəkildə araşdırılır, tələləri, dizayn təlimatlarını və mobil şifrələnmiş TC-də DL-nin açıq məsələləri vurğulanır [4].

Mobil Şifrələnmiş Trafiki Təhlükəsizlik/xidmət keyfiyyətinə tətbiqi. Trafik təsnifatı müxtəlif tətbiqləri, trafik növlərini müəyyən etmək üçün paketdən və ya axından alınan məlumatların təhlilini əhatə edən bir texnikadır.

Trafik təsnifatı xidmətin keyfiyyətinə zəmanət, şəbəkə resurslarından optimal istifadə, kənarə çıxmaların, zərərli proqram trafikinin və şəbəkə müdaxiləsinin aşkarlanması kimi şəbəkə idarəetmə tapşırıqlarında mühüm rol oynayır. Bu tapşırıqlar üçün dəqiq trafik təsnifatı alətləri lazımdır. İnternet istifadəçiləri üçün məxfilik və

təhlükəsizlik problemlərinin artması səbəbindən bir çox proqramlar müxtəlif növ trafik şifrələməsindən istifadə edir. Təhlükəsizlik/xidmət keyfiyyətinə nəzarət cihazları və şəbəkə monitorları kimi bir sıra alətlər trafiki yaradan proqram haqqında bilikləri fərz etməklə işləyir. Bu tələblər tam yerinə yetirilmədikdə məhdudlaşdırılır (və ya zədələnir).

Şəbəkə trafikinin xüsusi proqramlarla əlaqələndirilməsi prosesi Trafik Təsnifatı (TC) kimi tanınır və bir neçə sahədə çoxdan formalaşmış tətbiqə malikdir. TC ev və müəssisə şəbəkələri İnternet üzərindən hərəkət edən trafikin xarakterini dəyişdirən əl cihazlarının kütləvi yayılması ilə (İnternet istifadəsinin son qiymətləndirmələri tərəfindən dəstəklənir) getdikcə daha çox uzaqlaşır.

Beləliklə, mobil TC-nin həm zərurəti, həm çətinliyi, qiymətli profil məlumatı (məsələn, reklamçılar, sığorta şirkətləri və təhlükəsizlik agentlikləri üçün) potensialı ilə gücləndirilmiş (TC üçün ümumi sürücülərdən başqa), həm də məxfiliyi mənfi cəhətlər (məsələn, sağlamlıq və tanışlıq proqramları kimi kontekstə həssas tətbiqlərin tanınması və şirkətlərdən öz cihazınızı gətirin siyasətləri halında) nəzərdə tutulur [7].

TC, mobil trafik kontekstində hətta kəskinləşən, adətən ayrı-seçkilik edilməli olan çoxlu sayda proqram və hər bir tətbiq üçün kifayət qədər olmayan təlim nümunələri ilə xarakterizə olunan, qənaətbəxş performansa nail olmağa mane olan öz çətinlikləri və tələbləri ilə gəlir. Bundan əlavə, şifrələnmiş protokolların (TLS), eləcə də NAT və dinamik portların artan qəbulu Dərin Paket Təftişi (DPI) və port əsaslı metodlar kimi qurulmuş yanaşmaları məğlub edərək dəqiq dərəcə təsnifatı üçün yeni problemlər yaradır.

Həqiqətən də, Şifrələnmiş Trafikin (ET) mövcudluğu ciddi məhdudiyyətdir ki, onu yalnız qapalı dünya müəssisə ssenarilərində ortada man-in-the-in-the-mid proxies kimi həll yollarından istifadə etməklə keçmək olar. Beləliklə, ML əsaslanan təsnifatlar, xüsusən də bu kontekstdə ən uyğun hesab olunur, çünki onlar ET-yə də uyğun gəlir. Lakin mütləq port məlumatlarına etibar edilmir [10].

Mobil Şifrələnmiş Trafiki ilə domen-ekspert tərəfindən idarə olunan xüsusiyyətlərin əldə edilməsi. Standart ML təsnifatçılarının uğurlu istifadəsi TC kontekstində paketlərin ardıcıllığından və ya mesaj ölçülərindən çıxarılan statistikaya uyğun gələn əl işi (domen-ekspert tərəfindən idarə olunan) xüsusiyyətlərin əldə edilməsinə əsaslanır. Bu cür proses çox vaxt aparır, ancaq avtomatlaşdırmaya uyğun gəlmir.

Mobil trafikin təkamülü və qarışığı ilə müqayisədə sürətlə köhnəlir, daim hərəkətdə olan hədəf hesab olunur. Dəqiq və müasir mobil trafik təsnifatçılarının dizaynını "ənənəvi" ML yanaşmaları ilə istisna edir. Buna uyğun olaraq, biz hesab edirik ki, strukturlaşdırılmış xüsusiyyət təmsillərini avtomatik öyrənərək təsnifatlaşdırıcıları birbaşa giriş məlumatlarından öyrətməyə imkan verən Dərin Öyrənmə (DL) dinamik və çətin mobil TC-də yüksək performansa nail olmaq yolunda addım atmış ola bilər.

Bu tədqiqat işində, mobil TC ssenarisində azaldılmış DL texnikalarının müqayisəsi üçün sistematik çərçivə təmin etmək məqsədi daşıyır. Bu, TC ədəbiyyatında bu yaxınlarda ortaya çıxan və burada təkrarlanan bir neçə DL təsnifatının tənqidi təhlilindən irəli gəlir. Təfərrüatlı olaraq, təklif olunan çərçivə problemi müxtəlif nöqteyi-nəzərdən tədqiq edir, məsələn:

- ✓ qəbul edilmiş TC obyekti,
- ✓ DL təsnifatına verilən giriş məlumatlarının növü,
- ✓ istifadə edilən DL arxitekturası və
- ✓ tələb olunan performans tədbirləri toplusu [11].

Çərçivəmizin bir nümunəsi olaraq, ən cəlbedici texnikaları qiymətləndirmək və real vaxt, DL vasitəsilə dəqiq mobil TC üçün açıq məsələləri vurğulamaq üçün real insan istifadəçilərinin fəaliyyətinin üç verilənlər bazasına əsaslanan illüstrativ müqayisəni təqdim edirik. Bildiyimiz qədər, bu günə qədər mobil ssenaridə oxşar sistemli yanaşma və eksperimental araşdırma aparılmamışdır. Bu işin nəticələri

mövcud DL-əsaslı trafik təsnifatlarının çatışmazlıqlarını və aşağıdakılara olan ehtiyac olduğunu vurğulayır:

1. trafik məlumatlarından ekstrapolyasiya edilmiş qərəzsiz, informativ və heterojen girişlər,

2. daha mürəkkəb DL arxitekturaları və

3. ciddi performans qiymətləndirmə iş masası.

Qalan hissəsi is aşağıdakı kimi təşkil edilmişdir:

✓ ML əsaslı mobil TC-də mövcud ən müasir texnologiyaları və DL arxitekturalarını standart TC-yə tətbiq edən son işlərə, əsas aspektlərə diqqət yetirməklə, mobil TC üçün ümumi DL çərçivəsini təsvir edilir;

✓ performansın qiymətləndirilməsi üçün işçi masası təsvir edilmişdir;

✓ eksperimental nəticələr müzakirə edilir;

✓ nəhayət, mobil mesajlar təqdim edilir və açıq problemləri vurğulanır [13].

3.2. Dərin öyrənmə tətbiqi üçün təlim, doğrulama və çox tapşırıqlı öyrənmə prosesi.

Son illərdə akademik və sənaye ədəbiyyat icmaları nəqliyyatın təsnifatına çox diqqət yetirir. Trafik təsnifatı üçün port əsaslı və dərin paket yoxlaması (DPI) kimi ənənəvi üsullar şifrələnmiş trafikdə eksponensial artım səbəbindən artıq adekvat deyil.

Buna görə də, şifrələnmiş şəbəkə trafikinin müəyyən edilməsi, təsnifləşdirilməsində port əsaslı və DPI yanaşmalarının məhdudiyyətlərini həll etmək üçün maşın öyrənməsinə əsaslanan üsullar təklif edilmişdir.

Son illərdə, əl ilə xüsusiyyətlərin çıxarılması kimi ənənəvi maşın öyrənməsinin məhdudiyyətləri səbəbindən trafik təsnifatı üçün dərin öyrənmə diqqəti daha çox cəlb etmişdir. Bu bölmədə biz hər bir deep learning metodu üçün xülasəsi göstərilən bəzi yeni sənədlərindən olan təlim və doğrulama haqqında qısaca məlumat təqdim edirik [20].

Dərin öyrənmə tətbiqi üçün təlim və doğrulama addımı ən yaxşı dəqiqliyi əldə etmək üçün modelin hiperparametrlərinin tənzimləndiyi hər hansı digər DL proqramlarına bənzəyir. Tipik olaraq, verilənlər toplusu üç ayrı dəstə bölünür:

qatar,

doğrulama və

test dəsti.

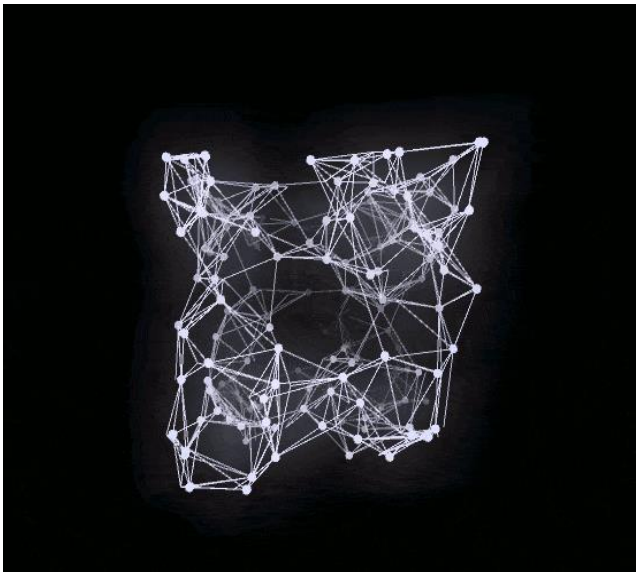
Model qatar dəstində öyrədilir və modelin hiperparametrlərini tənzimləmək üçün doğrulama dəstinin dəqiqliyi müşahidə edilir. Nəhayət, qərəzsiz dəqiqlik test dəstindən istifadə etməklə əldə edilir. Son iki addımın ən yaxşı ətraflı praktikaları bu tədqiqatın əhatə dairəsi xaricindədir. İstənilən digər tətbiqlərdə təlim və təsdiqləmə təlimatını oxuya və eyni ən yaxşı təcrübələri burada tətbiq edə bilərik [23].

Çox tapşırıqlı öyrənmə. Bu yanaşma birdən çox itki funksiyasının optimallaşdırıldığı istənilən modelə aiddir. Tipik bir yanaşma gizli təbəqələri bütün tapşırıqlar arasında bölüşdürməkdir, halbuki hər bir tapşırığın öz çıxış təbəqəsi var. Göstərilmişdir ki, o, həddindən artıq uyğunlaşma riskini azaldır və modelin müvafiq xüsusiyyətləri daha tez tapmasına kömək edir.

Bu, giriş məlumatları oxşar ehtimal paylanmasından yarandıqda və ya bir-birindən çevrilmələr toplusundan istifadə etməklə yaradıla bildikdə işləyir. Nəticə etibarilə, əlavə mövcud verilənlər toplusundan istifadə etmək və hər biri üçün bir tapşırıq müəyyən etmək mümkün ola bilər, əgər onlar hədəf tapşırıq verilənlər dəstinizə oxşardırsa. Bu, verilənlər bazasını asanlıqla artıra və ümumiləşdirməni yaxşılaşdıra bilər [29].

Çox tapşırıqlı öyrənmənin bir çox variantı təbii dilin işlənməsi və kompüter görmə üçün uğurla istifadə edilmişdir. Göstərilmişdir ki, hətta bir tapşırıqlı problemlər üçün bəzi köməkçi tapşırıqların əlavə edilməsi ümumiləşdirmə və performansı yaxşılaşdıracaqdır. Bununla belə, şəbəkə trafikinin təsnifatı tapşırığı üçün tədqiq edilməmişdir. Əlavə etiketləşdirməyə ehtiyac olmadan köməkçi tapşırığı müəyyən etməyin potensial olaraq bir çox yolu var.

Məsələn, ilk 20 paketin zaman seriyası məlumatını giriş kimi qəbul edən tipik bir modeli fərz edək. TCP/UDP sinifinin aşkarlanması, bütün axının orta paket uzunluğunun proqnozlaşdırılması, siçanlar/fil axınının aşkarlanması və s. kimi insan etiketinə ehtiyacı olmayan bir çox köməkçi tapşırıqlar müəyyən edilə bilər. Şəbəkə trafikinin təsnifatı üçün hələ də çox tapşırıqlı öyrənmənin effektivliyi öyrənilməmişdir [31].



Şək. 3.1. Şəbəkə trafikinin təsnifatı

3.3. Şifrələnmiş Trafikinin Sinifləndirməsi üçün Deep Learning tədqiq üsulları.

Xüsusiyyət mühəndisliyi məlumatların çıxarılması və seçim üsullarını əhatə edən trafik təsnifatının kritik aspektidir. Şifrələnmiş Trafikinin Sinifləndirməsi ən effektiv yol hərəkəti xüsusiyyətlərinin seçilməsi bir sıra amillərdən asılı olan yüksək təsnifat dəqiqliyinə nail olmaq üçün vacibdir. Xüsusiyyətlər həm hesablama mürəkkəbliyinə, həm də təsnifat üçün tələb olunan paket sayına (yaddaş mürəkkəbliyi), giriş strukturuna və ölçülərinə birbaşa təsir göstərir. Bundan əlavə, seçilmiş xüsusiyyətlərə əsasən uyğun model seçilməlidir. Müxtəlif tədqiqatçılar məkan-zaman, statistik və ya hibrid metodlardan istifadə edərək şifrələnmiş trafiki təsnif etmişdir. Bununla belə, şifrələnmiş trafik məlumatlarının unikal təbiətinə görə, bir xüsusiyyət növünə diqqət

yetirmək aşağı təsnifat dəqiqliyinə səbəb ola bilər. Dərin öyrənmə şəbəkələri, şifrələnmiş trafik axınlarından və ya seanslardan çıxarılan yüksək səviyyəli xüsusiyyətlərə əsaslanan dəqiq trafik təsnifatı təklif edir. Zaman, məkan və statistik xüsusiyyətlərin inteqrasiyasının üstünlüklərini başa düşmək üçün əvvəlcə onları aşağıdakı kimi müəyyən etmək lazımdır [32].

- Müvəqqəti xüsusiyyətlər: Paket tezliyi, ölçüsü və sessiya müddəti kimi müvəqqəti xüsusiyyətlər, nümunələri müəyyən etməyə və müvafiq olaraq təsnif etməyə kömək edən şifrələnmiş trafikin zamandan asılı xüsusiyyətləridir.

- Məkan xüsusiyyətləri: Məkan xüsusiyyətləri iş sessiyası ərzində şifrələnmiş trafikdə şəbəkə paket baytlarının yerini müəyyən etmək üçün çox vacibdir. Bu xüsusiyyətlər baytlar arasında məkan münasibətlərini ələ keçirir.

- Statistik xüsusiyyətlər: Orta paket ölçüsü, standart kənarlaşma və gəlişlər arası vaxt kimi statistik xüsusiyyətlər şifrələnmiş trafiki təsnif etmək və nümunələri müəyyən etmək üçün istifadə olunur. Bu effektiv təsnifat və təhlilə imkan verir [35].

Şifrələnmiş trafik təsnifatında müvəqqəti, məkan və statistik xüsusiyyətlərin birləşdirilməsi müxtəlif şifrələnmiş trafik növlərinə qarşı dəqiqliyi, səmərəliliyi və möhkəmliyi artırır. Bu üsul potensial təhlükəsizlik təhdidlərinin hərtərəfli təhlilinə və müəyyənləşdirilməsinə imkan verən trafik haqqında daha dərin məlumat verir. Şifrələnmiş trafiki dəqiq müəyyən etmək müxtəlifliyinə görə çətindir. Bununla belə, çoxsaylı funksiyalardan istifadə etməklə, daha geniş spektrli trafik xüsusiyyətlərini ələ keçirə bilər, bu da onu şifrələmə üsullarında dəyişikliklərə daha davamlı edir [80].

Bu yanaşma təsnifat nəticələrinin şərh oluna bilməsini yaxşılaşdırır, nəqliyyatın strukturu və davranışını hərtərəfli başa düşməyə imkan verir. O, çevik məkan, zaman və statistik xüsusiyyətlərə malikdir. Şifrələnmiş trafik təsnifatında çevik xüsusiyyətlər tətbiqdən və ya protokoldan asılı olmayaraq təsnifata imkan verir, trafikin mahiyyətinin hərtərəfli başa düşülməsini təmin edir. Bu bölmədə, aşağıdakı baxış nöqtələrinə diqqət yetirməklə TC-də ən müasir DL-ni tədqiq edilir:

1. trafik görünüşü (yəni, trafik aqreqatının növü),

2. DL-yə verilən giriş məlumatının növü və
3. istifadə edilən DL arxitekturası [38].

Qeyd etmək lazımdır ki, TC üçün təklif olunan bütün DL təsnifatçıları diqqətlə təhlil edilmiş və təkrar emal edilmişdir (məsələn, onların müvafiq işlərində təklif olunan hiper-parametr dəyərlərini təyin etməklə və ya sonuncular barədə məlumat verilmədikdə əsas tənzimləmə prosedurunu yerinə yetirməklə). Xüsusilə, aşağıda təsvir olunan yanaşmaları həm həyata, həm də sınaqdan keçirmək üçün TensorFlow üzərində işləyən Keras (Python) API tərəfindən hazırlanmış DL modellərindən istifadə edilir [41].

a) *Trafik Görünüşü.* TC ədəbiyyatında müxtəlif nəqliyyat obyektləri nəzərdən keçirilmişdir. Müəyyən bir trafik obyektinin tərifi xam trafikin çoxlu diskret trafik vahidlərinə necə bölündüyünü müəyyən edilir. Qeyd etmək lazımdır ki, DL istifadə edərək TC-yə yaxınlaşan bütün işlər istisna olmaqla, müvafiq təsnifat obyektləri kimi ya axınları, ya da iki axınları nəzərdə tutulur. Daha dəqiq desək, axın istiqamətləri nəzərə alınmaqla eyni 5 dəstli (yəni mənbə IP, mənbə port, təyinat IP, təyinat portu və nəqliyyat səviyyəli protokol) olan bütün paketlər kimi müəyyən edilir. Fərqli olaraq, iki axın müəyyən bir dəsti paylaşan trafikin hər iki istiqamətini əhatə edir (yəni mənbə və təyinat bir-birini əvəz edir).

Nəhayət, müvafiq təsnifat obyekti tək paketdir (yəni təsnifat proseduru paket səviyyəsində həyata keçirilir), TC problemi üçün ən yaxşı qranularlığa uyğundur (və faktiki olaraq müvafiq təsnifat tapşırığı üçün ən çətin quraşdırmanı təmsil edir).

b) *Daxil edilən verilənlərin növləri.* Tədqiq olunan DL arxitekturalarına verilən məlumatların tipini təxminən üç növə bölmək olar:

1. I TC obyektinin faydalı yükünün ilk N baytı;
2. II TC obyekti ilə əlaqəli PCAP faylına aid xam verilənlərin ilk N baytı;
3. I Np paketlərinin III informativ məlumat sahələri [44].

Birinci halda, DL arxitekturasına qidalanan məlumatlar ikili formatda giriş məlumatları ilə yalnız faydalı yüklə təmsil olunur. Bütün bu işlərdə faydalı yük bayt

şəklində düzülür və onu daxilində məhdudlaşdırmaq üçün normallaşdırılır (255 ilə). Seçim həmişə DL arxitekturası üçün giriş ölçüsünü azaltmaq üçün bir vasitə kimi əsaslandırılır. Digər tərəfdən, seçilən yükün təbəqəsi və ölçüsü konkret işdən asılıdır. Məsələn, bunlar TCP yükünün ilk 1000 baytına uyğundur. Oxşar seçim "L7" kimi etiketlənmiş giriş üçün edilir, burada TCP/IP modelində tətbiq səviyyəsindən 784 bayt nəzərə alınır. Fərqli şəkildə, müəlliflər ikinci qatdakı ilk 1500 faydalı yük baytını, yəni IP başlığını və hər bir IP yükünün ilk 1480 baytını nəzərdən keçirirlər ki, bu da 1500 bayt giriş vektoru ilə nəticələnir [53].

İkinci növ giriş məlumatı bütün protokol səviyyələrindən ("BÜTÜN" təbəqələri ilə qeyd olunur) məlumat toplamağa çalışır, çünki bəzi müvafiq hallarda VII səviyyədən aşağı səviyyələrdən olan məlumatlar da qeyd edildiyi kimi bəzi faydalı trafik məlumatlarını ehtiva edir (nəqliyyat qatı portları və ya bayraqlar kimi). Sonra, nəzərdən keçirilən məlumatlar adətən məlumat bağlantısı səviyyəsində tutulduğundan, ikinci səviyyənin çərçivələrindən faydalı yük çıxarılır. Bununla belə, bu halda təqdim edilən trafik həmişə təsnifat nəticələrində qərəzlilik yarada biləcək məlumatları ehtiva edən PCAP faylları şəklində olur. Xüsusilə, hər bir TC obyektinin yalnız ilk 784 baytı istifadə olunur [56].

Nəhayət, üçüncü növ daxiletmə məlumatı ilk Np paketlərinin seçilmiş protokol sahələri (şifrələnmiş faydalı yükün açıq şəkildə yoxlanılması ilə bağlı olmayan) ilə təmsil olunur. Məsələn, müəlliflər yalnız TC obyektinə (biflow) mübadilə edilən ilk 20 paketi nəzərə alır və hər paket üçün aşağıdakı 6 sahə çıxarılır (beləliklə, hər bir TC obyekti üçün 20×6 matris əldə edilir): mənbə və təyinat portları, nəqliyyat qatının faydalı yükündəki baytların sayı, TCP pəncərəsinin ölçüsü, daxil olma vaxtı və paket istiqaməti ($\in \{0, 1\}$). Biz onu da qeyd etməliyik ki, paket/mesaj istiqamətləri ardıcıllığı bu yaxınlarda DL-əsaslı vebsayt barmaq izində istifadə edilmişdir. Beləliklə, yuxarıda göstərilən bütün hallarda, hesablanmış sabit uzunluqlu məlumat girişlərindən daha uzun və ya daha qısa nümunələr ola biləcəyini qeyd edərək müzakirəni yekunlaşdırırıq. Belə hallarda, daha uzun nümunələr baytların və ya paketlərin nəzərdə tutulmuş

uzunluğuna qədər kəsilir, birinci/ikinci və ya üçüncü tip məlumatlarda, halbuki daha qısa nümunələrdə həmişə müzakirə olunan bütün mətnlərdə sıfırlarla doldurulma tətbiq edilir və işləyir [49].

c) *DL Təsnifat Alqoritmləri.* Burada TC üçün istifadə edilən DL arxitekturalarını nəzərdən keçiririk. Bu məqsədlə biz təlim dəstinin birinci girişini (M nümunəsindən hazırlanmış) x(i) kimi, müvafiq etiketi isə `(i) ilə təyin edirik. Bütün nəzərdən keçirilən DL təsnifatçıları, geri yayılma yolu ilə standart yerli optimallaşdırıcılar (məsələn, SGD, Adam və s.) tərəfindən əldə edilən kateqoriyalı çarpaz entropiyanı minimuma endirmək üçün öyrədilir [50].

3.4. Şifrələnmiş Trafik klassifikasiyası və naməlum məlumatların aşkarlanması üçün Dərin Öyrənmə metodları ilə əldə edilənlərin parametrlərin təhlili.

Texnologiyanın inkişafı bir çox səbəbə görə şifrələnmiş trafikin təhlilini tələb edir. Şifrələnmiş trafik zərərli fəaliyyətləri gizlədir, aşkarlama və təhlili daha çətinləşdirir. O, həmçinin kritik tətbiqlərin kifayət qədər bant genişliyi və prioritet almasını təmin etməklə şəbəkə performansını optimallaşdırır.

Şifrələnmiş trafikin təhlili tənzimləmə uyğunluğu üçün də çox vacibdir, çünki səhiyyə və maliyyə kimi sənayelərdə həssas məlumatlar məlumatların məxfiliyi qaydalarına uyğun olmaq üçün şifrələnməlidir. Trafik təsnifatı yanaşmaları ümumiyyətlə dörd kateqoriyaya bölünür:

- port əsaslı,
- faydalı yükə əsaslanan,
- maşın öyrənməsinə əsaslanan və
- dərin öyrənməyə əsaslanan [59].

Birinci və ən sadə yanaşma, port əsaslı təsnifat paket başlığından port nömrəsini çıxararaq trafikin növünü müəyyən edir. Dərin Paket Təftişi (DPI) kimi tanınan faydalı yük əsaslı analiz paket yüklərini müxtəlif protokollar üçün əvvəlcədən təyin edilmiş

nümunələrə və imzalara qarşı yoxlayır. Həm port əsaslı, həm də DPI yanaşmalarının müəyyən məhdudiyyətləri var. Porta əsaslanan yanaşma, portun yanlışlığı, təsadüfi port nömrələrinin təyin edilməsi, dinamik portların istifadəsi və şəbəkə ünvanının tərcüməsi (NAT) texnikası səbəbindən trafikin təsnifatında daha səmərəli ola bilər. Bundan əlavə, məxfiliyi qorumaq üçün şifrələmə üsulları internet trafikində geniş istifadə olunur. Nəticə etibarilə, paket yükündən faydalı məlumatların çıxarılması daha az effektivdir, bu da daha az dəqiq təsnifatla nəticələnir [62].

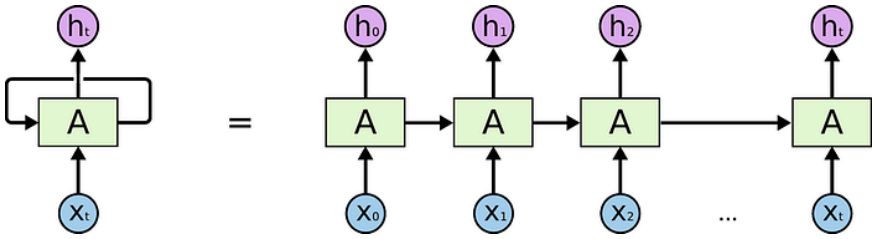
DPI metodu ilə bağlı əsas problem onun yüksək hesablama xərcləridir ki, bu da onu real vaxt və ya şifrələnmiş trafik təsnifatı üçün yararsız edir. Buna görə də, əvvəlki iki metodun məhdudiyyətlərini aradan qaldırmaq üçün statistik axın təhlili və maşın öyrənmə yanaşmaları tədqiq edilmişdir. ML alqoritmləri trafik məlumatlarında fərqləndirici xüsusiyyətləri, nümunələri öyrənmək üçün zaman sıralarından və statistik məlumatlardan istifadə edir.

Maşın öyrənməsinə əsaslanan metodlar ənənəvi yanaşmalarla problemləri həll edir. Lakin domen mütəxəssisləri tələb edən, vaxt apara bilən əl işi paketləri və sessiya xüsusiyyətlərini ələ keçirmək kimi yeni problemlər ortaya çıxarır. Başqa sözlə, maşın öyrənmə üsulları insan tərəfindən hazırlanmış xüsusiyyətlərdən çox asılıdır, bu da dəqiqliyi və ümumiləşdirməyi məhdudlaşdırır.

Maşın öyrənmə problemlərini həll etmək üçün yeni trafik təsnifatı metodu tələb olunur. Maşın öyrənmə problemlərini həll etmək üçün dərin öyrənmə yanaşmaları təklif edilmişdir. Klassik maşın öyrənmə alqoritmlərindən fərqli olaraq, dərin öyrənmə alqoritmləri funksiyaların çıxarılmasını avtomatik həyata keçirir və onları şifrələnmiş trafik təsnifatı üçün cəlbedici edir [67].

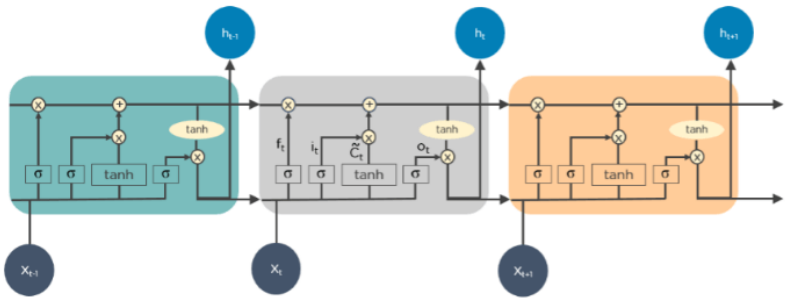
Dərin öyrənmə metodlarının digər üstünlüyü ondan ibarətdir ki, onlar daha mürəkkəb nümunələri ələ keçirir və nəticədə maşın öyrənmə metodlarından daha dəqiq təsnifatlar əldə edilir. Dərin öyrənmə, trafik axını məlumatlarından xüsusiyyətləri çıxarmaq üçün çox vaxt Konvolyusiya Neyron Şəbəkələrindən (CNN) istifadə edərək şifrələnmiş trafiki təsnif etmək üçün praktik bir üsuldur.

CNN-lər zamanla dəyişən dinamik trafik axını məlumatlarından xüsusiyyətləri çıxarmaq üçün həmişə ən yaxşı seçim olmaya bilər. Bu məlumatlar CNN-lərin tez-tez gözdən qaçırdığı iş sessiyalarının statistikasını ehtiva edir. LSTM və Bi-LSTM modelləri trafik axın məlumatlarından xüsusiyyətlər çıxara bilər. Məlumatların müvəqqəti xüsusiyyətlərini nəzərə almaq və trafikin təsnifatı tapşırığı üçün uyğun bir texnikadan istifadə etmək çox vacibdir. LSTM-ləri uzun müddət ərzində asılılıqları öyrənmək, uyğunlaşdırmaq üçün proqramlaşdırılmış RNN kimi təsvir etmək olar [79].



Şək. 3.2 Proqramlaşdırılmış RNN əməliyyatlarının sxemi

LSTM-lər yaddaşı, əvvəlki girişləri məhdudlaşdıra bildiyi üçün əsasən zaman sıralarının proqnozlaşdırılmasında istifadə olunur. Bir-biri ilə fərqli şəkildə əlaqə saxlayan 4 qarşılıqlı laydan ibarət zəncirvari quruluşundan irəli gəlir. Həmçinin, nitq tanıyıcılarının yaradılmasında, əczaçılıq sənayesinin inkişafında və musiqi dövrlərinin yaradılmasında da istifadə edilə bilər. LSTM hadisələr ardıcıllığı ilə işləyir. Həm müəyyən hüceyrə vəziyyəti dəyərlərini seçici şəkildə yeniləyir, həm də xüsusi hissələrini çıxış kimi istehsal edir [68].



Şək. 3.3 LSTM əməliyyatlarının diaqramı

CNN-lərin məhdudiyyətlərinə baxmayaraq, onlar hələ də trafik axını məlumatlarından xüsusiyyətləri çıxarmaq üçün istifadə edilə bilər. Dərin öyrənmə xüsusiyyətləri insan müdaxiləsi olmadan çıxarsa da, balanssız öyrənmə məlumatları bu yanaşmada zəif göstəricilərlə nəticələnə bilər. Yarı nəzarət edilən seçmə üsulları verilənlər balanssızlığını aradan qaldırmaq üçün xüsusi siniflərdə məhdud nümunə ölçüləri olan verilənlər dəstləri üçün sintetik nümunələr yaratmaq üçün istifadə olunur.

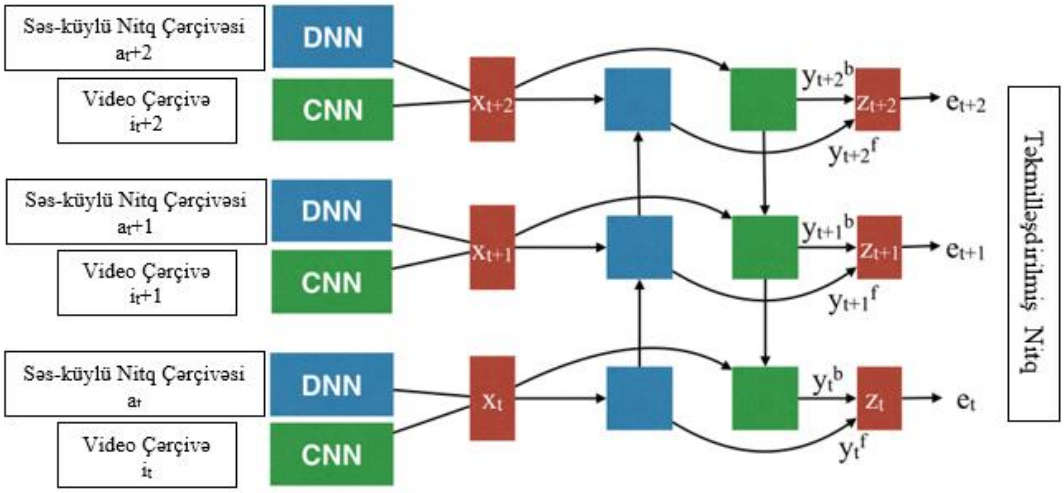
İnternet rabitəsində məxfiliyi təmin etmək üçün şifrələmə üsullarından geniş istifadə olunmasına baxmayaraq, mobil cihaz istifadəçiləri hələ də məxfilik və təhlükəsizlik risklərinə həssasdırlar.

Bu yazıda istifadəçi fəaliyyətinin aşkarlanması çərçivəsinə əsaslanan yeni Dərin Neyron Şəbəkəsi (DNN) işlənmiş şifrəli İnternet trafik axınından mobil tətbiqlərdə (tətbiqdaxili fəaliyyətlər kimi tanınır) yerinə yetirilən təfərrüatlı istifadəçi fəaliyyətlərini müəyyən etmək üçün təklif olunur. Çətinliklərdən biri odur ki, saysız-hesabsız tətbiqlər mövcuddur və onlardan bütün mümkün məlumatlardan istifadə edərək DNN modelini toplamaq və öyrətmək demək olar ki, mümkün deyil [71].

Bu motivasiya ilə bu tədqiqat işində biz DNN ilə şifrələnmiş trafik şəraitində tətbiqdaxili fəaliyyətləri müəyyən etmək üçün yeni çərçivə təklif edirik. İstifadəçilər tərəfindən mobil tətbiqlərdə həyata keçirilən fəaliyyətlərin müəyyən edilməsi mobil istifadəçinin vərdişlərini profilləşdirmək üçün istifadə edilə bilər. Bu, şəbəkələr daxilində və ümumilikdə marketinq və ya kəşfiyyat məqsədləri üçün istifadəçi kəşfi üçün faydalıdır. Təklif olunan metod DL-ə, təmsilçilik öyrənmə metoduna əsaslanır. Tədqiqatda əsas töhfələri aşağıdakı kimi ümumiləşdirilmişdir [73].

Sosial şəbəkələrdən tutmuş həyat tərzi, oyunlar, əyləncə, sağlamlıq, təhsil, maliyyə və s. Var olduğu qədər müxtəlif tətbiq kateqoriyaları var. Bu qədər geniş diapazon və bu tətbiqlərdəki hər bir tətbiqdaxili fəaliyyət üçün ML alqoritmlərini öyrətmək qeyri-mümkündür. Tətbiqdaxili fəaliyyət təsnifatı çərçivəsini real dünya mühitində yerləşdirərkən, çərçivə əvvəllər öyrədilməmiş fəaliyyətləri naməlum

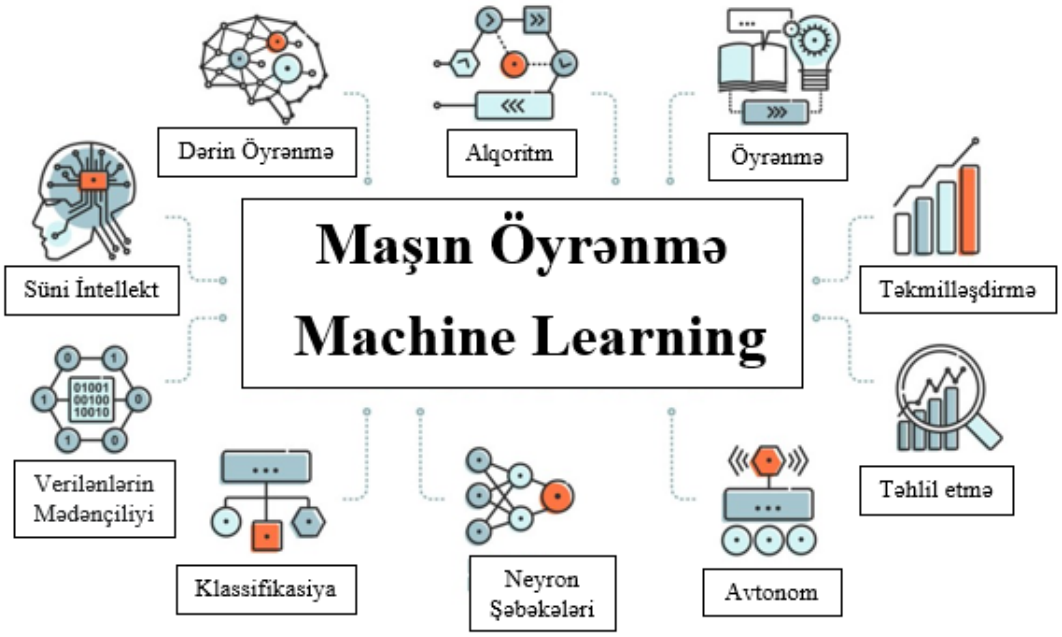
məlumat trafiki kimi təcrid edərkən həmin proqramlar daxilində tətbiqlər və fəaliyyətlər toplusunu müəyyən etməlidir [74].



Şək. 3.4 Multimodal hibrid dərin neyron şəbəkəsinin arxitekturası.

Mövcud ədəbiyyatların əksəriyyətində ML alqoritmləri eyni proqramlar toplusunda öyrədilir və sınaqdan keçirilir; bu, onları əvvəllər məlum olmayan trafikin filtrasiyası üçün yararsız edir.

Tədqiqat işində, təklif olunan çərçivə naməlum trafik tərəfindən yaranan səs-küyün mövcudluğunda şəbəkə trafikinin təhlilini dəqiq idarə etməyə qadirdir və beləliklə, bunu az və ya heç bir təlim olmadan yeni mühitlərə və məlumat axınlarına uyğunlaşa bilən arzuolunan bir yanaşma halına gətirir. Tətbiqdaxili fəaliyyətləri aşkar etmək cəhdində, dinləyicinin bütün tranzaksiya deyil, qismən şəbəkə trafikini tutması halları ola bilər, çünki istifadəçi fəaliyyəti artıq davam edə bilər. Bu hallarda, ədəbiyyatda mövcud olan əksər tədqiqatlar, onun imzası çəkilmiş pəncərəyə düşmürsə, hadisəni aşkar edə bilmir. Təklif olunan çərçivə hətta fəaliyyət trafikinin bir hissəsini müşahidə etməklə təfərrüatlı tətbiqdaxili fəaliyyətləri müəyyən etməyə qadirdir [76].



Şək. 3.5 Maşın Öyrənmə alqoritmləri.

Cari işlər ya gözdən keçirilir, və ya Instagram-da paylaşma, WhatsApp-da mesajlaşma və s. kimi qaba dənəli fəaliyyətlər yüklənir və s. Bu kimi ümumi fəaliyyətlərə diqqət yetirilir. Tədqiqatımız şifrələnmiş şəbəkə trafikində tətbiqdaxili istifadəçi fəaliyyətlərini təcrübi şəkildə müəyyən etməklə ən müasir vəziyyəti inkişaf etdirilir. Beləliklə, "mesaj göndərmə" kimi ümumi bir WhatsApp fəaliyyəti nəzərə alındıqda, bu mesajın uzun mətn, qısa mətn, şəkil, video və ya səs yazısı olduğu müəyyən edilə bilər. Şifrələnmiş domendə metadata istifadə edilərkən bu səviyyədə təsnifat çətindir. Çünki bu, dərin trafik nümunəsinin yoxlanılmasını tələb edir. Bununla belə, o, məxfi məlumatların saxlandığı istifadəçiləri müəyyən etmək üçün analitik üçün dəyərli məlumat verir. Təklif olunan çərçivə səkkiz müxtəlif tətbiqdən 92 tətbiqdaxili hadisəni müəyyən edə bilir. Facebook, Instagram, WhatsApp, Viber, Messenger, Gmail, Skype, YouTube kimi proqramlarda silsilə əməliyyatlar həyata keçirilərək hərtərəfli məlumat toplusu yaradılıb. Məlumat dəsti yeni tədqiqatları təşviq etmək və təqdim olunan nəticələrin təkrarlanmasına imkan yaratmaq üçün tədqiqat ictimaiyyəti ilə açıq şəkildə paylaşılır [77].

FƏSİL IV. Şifrələnmiş trafikin klassifikasiya üçün Deep learning metodlarının tətbiqi

4.1. Klassifikasiyada Deep Learning üçün tədbiqatlar və avantajlar

Şifrlənmiş trafikin klassifikasiyasında Deep Learning tətbiqi, bir çox avantaja malikdir. Bu metod, məlumatların dərin strukturlarını təhlil etmək üçün effektiv alqoritmlər təmin edir. İstifadə olunan tədbirlər və bu alqoritmlərin klassifikasiyada üstünlükləri:

- Avantajlar:

Ətraflı Məlumat Təhlili: Deep Learning, şifrəli trafik datalarini ətraflı şəkildə təhlil edərək, informasiya analizi sahəsində yeni bir dönəmin başlanmasına imkan verir. Bu, sinifləndirmə prosesinin datalari daha aydın şəkildə təhlil etməsinə kömək edir. Bununla yanaşı, gizli təhlükələri və doğru olmayan əlaqələri aşkar etməyə və artırılmış təhlükəsizlik tədbirlərini tanımağa kömək edir.

Böyük Məlumat Həcminin İşlənməsi: Gələcəkdə, İnternet bağlantılı cihazlar və multimedia tətbiqləri ilə əlaqələndirilmiş data həcmində böyük bir artım gözlənilir. Cisco'nun " Vizual Şəbəkə İndeksi" araşdırmasına görə, bu data həcmi 2023-ə qədər illik olaraq 5,3 zettabayta (5,3 trilyon gigabayt) çata bilər. Deep Learning alqoritmləri, bu böyük data həcmini effektiv bir şəkildə təhlil edərək, datalarin ən əhəmiyyətli hissələrini müəyyənləşdirə bilər. Bu da datalarin daha məqbul və anlamlı şəkildə işlənə biləcəyini göstərir.

Yeni Anlayışın İnkişafı: Deep Learning təlimi, data analizi sahəsində yeni bir dönəmin başlanmasına imkan verir. Bu da siber təhlükəsizlik sahəsində yeni və dərin bir anlayışın inkişafına kömək edə bilər. Deep Learning-in təmin etdiyi güclü analitik imkanlar, datalar arasında müxtəlif və effektiv davranışları daha yaxşı ayırt etməyə kömək edir. Bu faydalar, Deep Learning'in şifrəli trafik sinifləndirmə tətbiqinin əhəmiyyətini və effektivliyini vurğulayır. Bu metodun siber təhlükəsizlik sahəsində

əhəmiyyətli bir vasitə ola biləcəyi və gələcəkdəki data analizi ehtiyaclarını qarşılamaq üçün daha da geniş yayıla biləcəyi gözlənilir.

- Çoxqatlı Xüsusiyyətlər Təyin Etmə:

Şifrələnmiş trafikin klassifikasiyasında Deep Learning tətbiq etmək bir sıra avantaja malikdir. Bu, datalarin kompleks strukturlarını və dərinliklərini anlamağı təmin edən güclü alqoritmlər sayəsində mümkündür. Bu avantajlardan bəziləri:

Gələcək Proqnozlaşdırma. İnternetə bağlı cihazlar və multimedia tətbiqləri ilə əlaqələndirilmiş məlumatların həcmindəki artım, gələcəkdəki trendlərin proqnozlaşdırılmasını çətinləşdirir. Deep Learning alqoritmləri, böyük məlumat həcmindəki analizlər sayəsində gələcəkdəki meylləri müəyyənləşdirməyə kömək edə bilər. Bu, siber təhlükəsizlik strategiyalarının daha effektiv bir şəkildə planlanmasına imkan verir.

Fərqli Fəaliyyətlərin Tanınması. Deep Learning alqoritmləri, normal datalarla müqayisədə anormal fəaliyyətləri təyin etmək üçün müxtəlif xüsusiyyətləri qiymətləndirir. Bu, normal təhlükəsizlik qaydalarından ayrılan və potensial təhlükələri tanımaq üçün önəmlidir. Məsələn, icazəsiz və ya gözlənilməyən bir şəbəkə əlaqəsi istifadə edərək məlumat göndərən bir cihazın tanınması [8].

- Öz-Öyrənmə və Öz-Adaptasiya

Şifrələnmiş trafikin təsnifatında Deep Learning, öz-öyrənmə və öz-adaptasiyanın əhəmiyyətini dəyərləndirir. Bu, sistemlərin yeni məlumatlar öyrəndikcə və dəyişən mühitlərə uyğunlaşdıqca effektivliyini artırır.

Öz-Öyrənmə. Deep Learning alqoritmləri, məlumatlarla təmin olunduqda, mövcud nümunələrdən öyrənmə və məlumatların öz-öyrənmə prosesini tətbiq edir. Bu, sistemin yeni və müstəqil məlumatlardan öyrənməsinə kömək edir və istifadəçi tərəfindən əlavə məlumatlara ehtiyac olmadan təhlükəsizlik tədqiqatını davam etdirməsinə imkan verir. Məsələn, yeni vəziyyətlərə uyğun protokolların öz-öyrənməsinə əsaslanan şəbəkə əməliyyatlarının təyin edilməsi.

Öz-Adaptasiya. Deep Learning alqoritmləri, dəyişən mühit şəraitlərinə uyğunlaşmağa imkan verən öz-adaptasiya mexanizmləri ilə təchiz edilmişdir. Bu, sistemlərin fəaliyyətini dəyişən şəraitlərə uyğunlaşdırmaq və effektivliyi artırmaq üçün əhəmiyyətli bir vasitədir. Məsələn, trafik məlumatlarında gözələnən dəyişikliklərə cavab verən adaptiv sinifləndirmə modelləri. Bu mexanizmlər, əvvəlcədən proqramlaşdırılmış qaydaların və parametrlərin dəyişdirilməsini tələb etmədən, sistemin yeni məlumatlar və şəraitlərə avtomatik olaraq uyğunlaşmasına imkan verir. Bu da məlumatları effektiv bir şəkildə təhlil etməsinə və təhlükəli fəaliyyətləri müəyyənləşdirməsinə imkan verir.

- Nümunələr:

Bir Deep Learning modeli, şəbəkədəki fəaliyyətləri təhlil edərək, hansı məlumatların normal və hansılarının anormal olduğunu müəyyənləşdirir. Bu proses zamanla dəyişən təhlükəli fəaliyyətləri tanımaq üçün sistemə əlavə məlumatlar əlavə etmədən öz-öyrənmə imkanı verir. Bu, sistemlərin özünəmənən öyrənmə mexanizmləri ilə məlumatların dəyişən təhlükəli fəaliyyətlərini müəyyənləşdirməsinə və buna uyğun təhlükəsizlik tədbirləri götürməsinə kömək edir. Sistem, potensial bir hücum növünü öz-öyrənərək, müdafiə strategiyalarını avtomatik olaraq tənzimləyə bilər. Məsələn, bir hücum növünü tanımaq və avtomatik olaraq müdafiə siqnalları göndərmək üçün şəbəkədəki yeni hücum növlərini öz-öyrənən bir sistem, sistemə təhlükəli fəaliyyətlərə qarşı daha sürətli və daha effektiv bir reaksiya vermə imkanı verir [19].

- Yüksək Dəqiqlik və Performans

Yüksək Dəqiqlik. Dünya İnternet Təhlükəsizliyi Koordinasiya Mərkəzi (ISC) tərəfindən aparılan bir araşdırma, şifrlənmiş trafik həcmində artımın cəmiyyət tərəfindən görünməyən təhlükələrin artmasına səbəb olduğunu göstərir. Deep Learning tətbiqatları, bu artan həcmi effektiv bir şəkildə təhlil etmək və təhlükəli fəaliyyətləri müəyyən etmək üçün dəqiq alqoritmlər təmin edir. 2020-ci ildə, Cisco'nun Təhlükəsizlik İnternet İndeksi (CSI) araşdırması, İnternetin ən azı 71%-inin (daha çoxu da mümkündür) artıq şifrələndirildiyini göstərir. Bu, klassifikasiya sistemlərinin daha

da kompleksləşdirilməsini və yüksək dəqiqlik tələb etməsini təmin edir. Deep Learning modellərinin sürəti və dəqiqliyi, hər gün milyonlarla yeni məlumatın təhlil edilməsini mümkün edir. Bu, potensial təhlükələrin tez bir şəkildə müəyyən edilməsini və əhəmiyyətli reaksiya vaxtının azaldılmasını təmin edir.

Yüksək Performans. Gələn ilə aid Prognosis Market Research tədqiqatına görə, şifrlənmiş trafikdə olan artım, təhlükəsizlik tədqiqatının sektorunun yalnız 2025-ci ilədək illik ortalama büxtə qurğusunda 2021-ci ildən 9.68% artım göstərəcəyini göstərir. Bu, yüksək performanslı təlim alqoritmlərinin tələbini artırır.

- Statistikalarla Təsdiq:

Dünya üzrə məlumat sənayesinin inkişafı ilə bağlı IDC tədqiqatına görə, hər saniyə ortalama olaraq 3.7 megabayt məlumat yaranır. Bu artan məlumat həcmi, klassifikasiya üçün daha güclü Deep Learning təlim alqoritmlərinin tələbini artırır. Gündəlik həyatımızda, sürücülərin avtomobillərinə qoyduğu sensör çipdən, sosial media platformalarında paylaşılan məlumatlara, istehsal prosesləri sənədlərindən, büdcələrə qədər bir çox sahədə məlumatlar yaranır. Bu məlumatlar arasında, potensial təhlükələri aşkar etmək və effektiv təhlil etmək üçün Deep Learning təliminin tətbiqi daha da əhəmiyyətli hala gəlir.

PwC tədqiqatına görə, kibertəhlükəsizlik təşkilatlarının 74%-i Deep Learningin effektiv klassifikasiya üçün əhəmiyyətini dəstəklədiyini göstərir. Bu, məlumatları təhlil etmə və təhlükəli fəaliyyətləri tanımaq üçün bu texnologiyaya olan marağın artmasını təsdiqləyir.

McAfee tədqiqatına görə, məlumatların dərin öyrənmə ilə təhlilinin, kibertəhlükəsizlik sahəsində ən optimal yolla ən dərin təhlükələri aşmaq üçün əhəmiyyətli bir yoldur. Bu, dərin öyrənmənin klassifikasiya üçün üstünlüklərini təsdiqləyir və bu sahədə tətbiq olunmasının əhəmiyyətini göstərir. Bu statistikalar, Deep Learning təliminin klassifikasiya üçün tətbiqinin əhəmiyyətini və effektivliyini təsdiqləyir və tədqiqatçıların və endustri işçilərinin bu texnologiyaya olan maraqlarını və inamını nümayiş etdirir [27].

4.2. İstifadə olunan Deep Learning alqoritmləri Şifrələnmiş Trafikdə Uyğunluq.

1. MLP: Şifrələnmiş trafikin meta məlumatlarını (Şifrələnmiş trafikin meta məlumatları dedikdə, şifrələnmiş məlumat paketlərinin məzmununu açmadan onlar haqqında bəzi əlavə məlumatlar nəzərdə tutulur.) təhlil etmək üçün uyğundur. Məlumatların non-lineer əlaqələrini öyrənmək üçün yaxşı performans göstərir, lakin şəkillər və zaman sıralı məlumatlar üzərində məhdudiyyətləri var.

2. CNN: Şifrələnmiş trafikin vizual aspektlərini, xüsusən də şifrələnmiş məlumatlardan çıxarılan görüntüləri və ya vizual məlumatları təhlil etmək üçün uyğundur. Lokal xüsusiyyətləri təsbit etmək üçün ideal olsa da, zaman sıralı məlumatlarda çətinlik çəkir.

3. RNN: Zaman sıralı şifrələnmiş trafikin təhlili və potensial təhlükəli davranışların aşkarlanması üçün idealdır. Keçmiş məlumatlarla işləmək qabiliyyəti RNN-ləri dəyişkən uzunluqlu trafikin təhlili üçün uyğun edir. Bu müqayisə göstərir ki, şifrələnmiş trafikin analizi üçün fərqli dərin öyrənmə arxitekturalarını tətbiq etmək, hər bir arxitekturanın spesifik güclü və zəif tərəflərini nəzərə alaraq, daha yaxşı nəticələr əldə etməyə imkan verə bilər. Optimal nəticələr üçün bu alqoritmlərin kombinasiyasını və ya hibrid modelləri istifadə etmək də faydalı ola bilər.

Cədvəl 4.1.

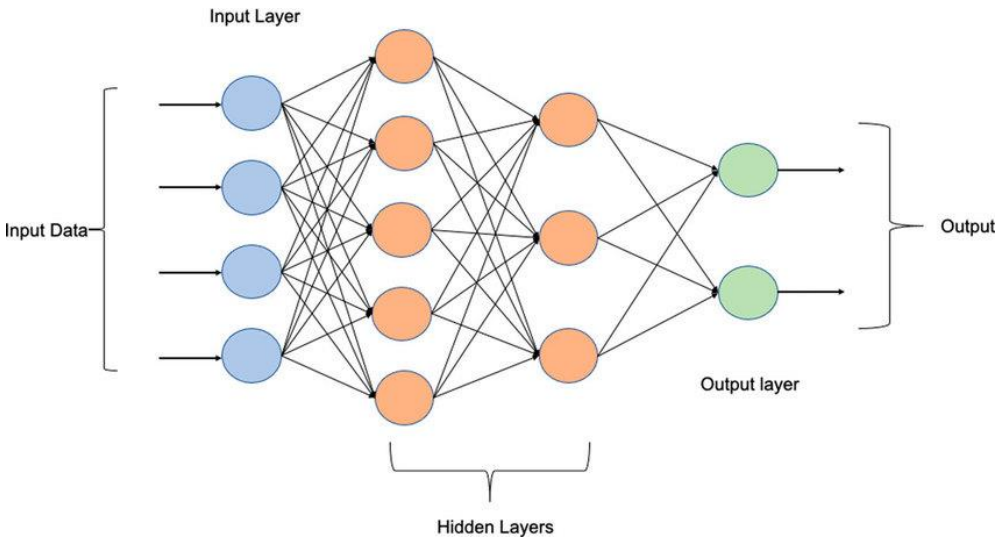
Optimal nəticələr üçün bu alqoritmlərin kombinasiyasını və ya hibrid modelləri

Alqoritm	**Üstünlüklər**	**Zəif Tərəflər**	**Şifrələnmiş Trafikdə Uyğunluğu**
MLP	Sadə və çevik, lineer olmayan əlaqələri öyrənə bilir	Məkanlı məlumatlarla zəif performans, böyük ölçülü məlumatlarla məhdud	Meta məlumatların təhlili üçün uyğundur
CNN	Məkanlı məlumatların təhlili, filtrlər və konvolusiya qatları	Uzunmüddətli ardıcıllıq əlaqələri ilə çətinliklər, yüksək hesablama yükü	Şəkillər və video məlumatları üçün uyğundur
RNN	Zaman sıralı məlumatların təhlili, dəyişkən uzunluqlu məlumatlar	Uzunmüddətli əlaqələrin öyrənilməsində çətinliklər, yaddaş tələbləri	Zaman sıralı şifrələnmiş trafik üçün uyğundur

Performansın Müqayisəsi

- Multi-Layer Perceptron (MLP)

MLP, Deep Learning alqoritmlərinin ən müasir və ən çox yayılan alqoritmlərindən biridir. Bu alqoritmin əsas prinsipi, bir neçə dərəcəli gizli layerlardan təşkil olunan bir neyron şəbəkəsində məlumatların işlənməsidir. MLP, hər bir layerdəki neyronların bir-biri ilə bərabər hərəkət etdiyi qatları istifadə edir. Hər bir neyron, əvvəlki layerdəki bütün neyronlara əlaqələr ilə bağlanır və bu, məlumatların non-lineer əlaqələrinin öyrənilməsinə imkan verir. Bu, şifrlənmiş trafikdəki mürəkkəb informasiyanın təhlil edilməsində və anlaşılmasında çox faydalıdır. MLP, məlumatları möhkəmləndirən və yoxlama məlumatlarına görə sinifləndirmə əməliyyatlarını yerinə yetirən müasir bir alqoritmdir. Əsasən, şifrlənmiş trafikin klassifikasiyasında MLP istifadə olunur. Bu alqoritm, məlumatların təhlil edilməsi, möhkəmləndirilməsi və sinifləndirməsi üçün əhəmiyyətli bir vasitədir. Əlavə olaraq, MLP tətbiqi nəzəriyyələri tədqiq etmək, məlumatların dərin analizini təmin etmək və şəbəkə təhlükələrinin aşkarlanmasına kömək edən müasir kibertəhlükəsizlik sistemləri üçün əhəmiyyətli bir texnoloji əsas təşkil edir [36].



Şək. 4.1 Multi-Layer Perceptron (MLP) alqoritmi.

Əgər MLP (Multi-Layer Perceptron) alqoritminin özünə aid bir model qurmaq istəyirsinizsə, Python vasitəsilə TensorFlow, PyTorch və ya Keras kimi dərin öyrənmə kitabxanalarından istifadə edərək bunu asanlıqla həyata keçirə bilərik. Aşağıda, MLP alqoritminin qurulması üçün geniş bir yol verərik:

1. **Kitabxana İmportu:** İstifadə etmək istədiyiniz dərin öyrənmə kitabxanasını (TensorFlow, PyTorch, Keras və s.) Python proqramımıza import edir.

```
import tensorflow as tf  # TensorFlow
from tensorflow.keras import layers, models  # Keras
```

2. **Məlumatların Hazırlanması:** MLP alqoritminin təlim edilməsi üçün uyğun məlumatlar hazırlanmalıdır. Bu məlumatlar çox zaman bir xarici mənbədən (məsələn, CSV faylı) yüklənir və praqmatik şəkildə təlim və sınaq məlumatlarına bölünür.

3. **Modelin Qurulması:** MLP alqoritminin arxitekturasını təyin etmək üçün bir model qurmaq üçün istifadə olunur.

```
model = models.Sequential([
    layers.Dense(128, activation='relu', input_shape=(input_size,)),  # İlk gizli qat
    layers.Dense(64, activation='relu'),  # İkinci gizli qat
    layers.Dense(num_classes, activation='softmax')  # Çıxış qatı
])
```

Bu, iki gizli qat və bir çıxış qatı olan sadə bir MLP modelidir. Biz hər bir qatın növünü, neyron sayını və aktivasiya funksiyasını istədiyimiz kimi tənzimləyə bilərik.

4.Modelin Təlim Edilməsi: Modeli hazırlanan məlumatlar ilə təlim etdikdə

```
model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
model.fit(train_images, train_labels, epochs=10, validation_data=(test_images, test_labels))
```

Burada, optimizer, itkilərin hesablanması üçün istifadə edilən alqoritm (SGD, və s.) seçilir. loss funksiyası modelin itkilərini ölçmək üçün təyin edilir.

5. **Modelin Sınaqdan Keçirilməsi:** Modeli sınaq məlumatları ilə sınaqdan keçirmək və performansını ölçmək üçün istifadə olunur.

test_loss, test_acc = model.evaluate(test_images, test_labels)

print('Test accuracy:', test_acc)

Bu prosesləri izləyərək MLP alqoritminin sadə bir modelini qurarıq.

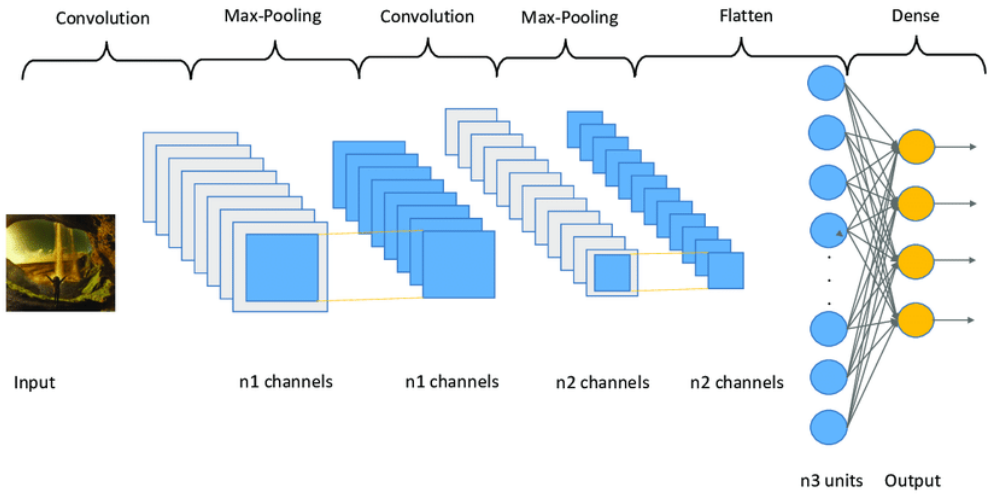
Aşağıdakı kod, TensorFlow-dan CIFAR-10 datasetini yükləyəcək, MLP modelini quraraq təlim edəcəyik:

```
import tensorflow as tf
from tensorflow.keras import layers, models

# CIFAR-10 datasetini yükləmək və bölmək üçün funksiya
def load_cifar10_data():
    (train_images, train_labels), (test_images, test_labels) = tf.keras.datasets.cifa

    # 0-1 aralığında normalizasiya etmək
    train_images, test_images = train_images / 255.0, test_images / 255.0

    return (train_images, train_labels), (test_images, test_labels)

# MLP modelinin qurulması
def build_mlp_model(input_shape, num_classes):
    model = models.Sequential([
        layers.Flatten(input_shape=input_shape),  # Düz bərabər sıxıştırma qatı
        layers.Dense(512, activation='relu'),     # İlk gizli qat
        layers.Dense(256, activation='relu'),     # İkinci gizli qat
        layers.Dense(num_classes, activation='softmax')  # Çıxış qatı
    ])
    return model

# CIFAR-10 datasetini yükləyin
(train_images, train_labels), (test_images, test_labels) = load_cifar10_data()

# Modeli qurun
input_shape = train_images[0].shape
num_classes = len(set(train_labels))
model = build_mlp_model(input_shape, num_classes)

# Modeli təlim edin
model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

model.fit(train_images, train_labels, epochs=10, validation_data=(test_images, test_l

# Modelin performansını yoxlayın
test_loss, test_acc = model.evaluate(test_images, test_labels)
print('Test accuracy:', test_acc)
```

Bu kod, CIFAR-10 ("Canadian Institute for Advanced Research" (Kanada İnkişafetmiş Tədqiqat İnstitutu) tərəfindən yaradılmış olan bir görüntü datasetidir.) datasetini yükləyəcək, daha geniş bir MLP modeli quracaq, onu təlim edəcək və sınaq məlumatları ilə performansını ölçəcək. Daha sonra, modeli optimallaşdırmaq, hiperparametrləri dəyişdirmək və performansını artırmaq üçün əlavə dəyişikliklər edə bilərik [45].

- Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN), şəkillər və video məlumatlarının təhlilində effektiv tətbiq olunur. Şifrlənmiş trafikin klassifikasiyasında, ən çox məlumatın vizual hissələri üçün CNN alqoritmləri istifadə olunur. CNN, şəkillər və video məlumatlarının təhlilində effektiv tətbiq olunur. Şifrlənmiş trafikin klassifikasiyasında, ən çox məlumatın vizual hissələri üçün CNN alqoritmləri istifadə olunur. Bu alqoritmlər, məlumatın lokal ətrafındakı patternləri təşhis etmək üçün dizayn edilmişdir. CNN alqoritmləri, məlumatın hər bir hissəsində müstəqil əlaqələr təşhis etmək üçün fərqli filtrlərdən istifadə edir. Bu filtrlər, məlumatın müxtəlif xüsusiyyətlərini aşkar etmək və mövcud patternləri təhlil etmək üçün öyrənilir. Məlumatın müxtəlif xüsusiyyətlərini öyrənir və təhlükəsizlik təhlilini yerinə yetirərkən məlumatın vizual hissələrində kritik xüsusiyyətləri aşkar etməyə kömək edir.



Şək. 4.2 Convolutional Neural Networks (CNN) alqoritmi.

CNN, klassifikasiya və təsnifat proseslərini yerinə yetirən kibertəhlükəsizlik sistemləri üçün əhəmiyyətli bir vasitədir. Bu alqoritmlər, şifrlənmiş trafik məlumatlarının təhlilində möhkəmləndirilmiş təhlükəsizlik strategiyalarının inkişaf etdirilməsinə kömək edir. CNN-in məlumat işləmə təcrübəsi, şifrlənmiş trafikin təhlilindəki dəqiqliyi və performansı artırır. CNN-in tətbiqi, potensial təhlükələri müəyyənləşdirərək şəbəkə təhlükəsizliyini artırmaq və mövcud təhlükələri aşkarlamaq üçün əhəmiyyətli bir vasitədir.

- Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNN), əsasən zaman sıralı datalarla çalışarkən təsirlidir. Bu algoritmalar, hər bir daxil olanlar sırasında əvvəlki addımlardan gələn məlumatlara əsasən çıxışlar edirlər. Şifrələnmiş trafikdə, RNN'lər çoxu zaman keçmiş trafiki və potansial olaraq təhlükəli davranışları analiz etmək üçün istifadə olunur.
RNN'lerin bir avantajı, her addımda fərqli uzunluqlardakı daxil olan datalarla işləyə bilməlidirlər. Bu, şifrələnmiş trafikdə dəyişən data uzunluqlarıyla başa çıxmak üçün idealdir.

Uzun zaman aralıqlarında və məhdud yaddaş problemlərində yönümlü olmaları bu modelin uzun vaxt əlaqələrinin öyrənilməsini məhdudlaşdırmaq və daha uzun məlumatlar üzərində doğru proqnozlar vermələrini çətinləşdirməyə səbəb olur. Bu səbəblə, daha kompleks və geniş məlumat setlərində işlərkən, RNN'lərin performansı aşağı düşə bilər.

- Autoencoders (AE)

Autoencoder'lar, əsasən nəzarətsiz öyrənmə üçün istifadə olunur, yəni təlim verilən məlumatlar üzərində işarələnmiş çıxışlar tələb olunmur. Bu xüsusiyyəti sayəsində, şifrələnmiş trafik kimi işarələnməmiş data toplusunda istifadə oluna bilər. Xüsusən, normal trafiki öyrənib anormallıqları aşkar etmək üçün istifadə oluna bilərlər. Bunun üçün normal trafikdə istifadə olunan autoencoder təlimi və sonra modelin bu normal trafikə nə qədər yaxşı uyğunlaşdığını qiymətləndirmək üçün anormallıqların aşkar edilməsi lazımdır.

Autoencoder'ların istifadə sahələri data sıxışdırma və kodlamadır. Şifrələnmiş trafikdə, trafiki daha az yer tutacaq şəkildə kodlamaq və saxlamaq üçün istifadə oluna bilərlər. Bu, saxlama məsələlərini azalda bilər və data ötürməsi üçün bant enişini optimize edə bilər.

- Generative Adversarial Networks (GAN)

Generativ Adversarial Networks (GAN) adı verilən alqoritma, son illərdə geniş tədqiqat və tətbiq sahələrində böyük marağa səbəb olmuşdur. Bu alqoritma, iki əsas modelin, yaradıcı (generator) və ayırıcı (discriminator), bir-birilə qarşı mübarizəsinə əsaslanır. Generator məlumatları təşkil edir, hər bir zaman orijinal verilənlərə uyğunluğu artırmaq üçün özünü təkmilləşdirir. Diqqət çəkən bir faktor odur ki, discriminator da eyni anda özünü təkmilləşdirir və bu, alqoritmanın effektivliyini artırır. Statistikalar göstərir ki, GAN-ların kibertəhlükəsizlik sahəsindəki rolu və önəmi artmaqdadır.

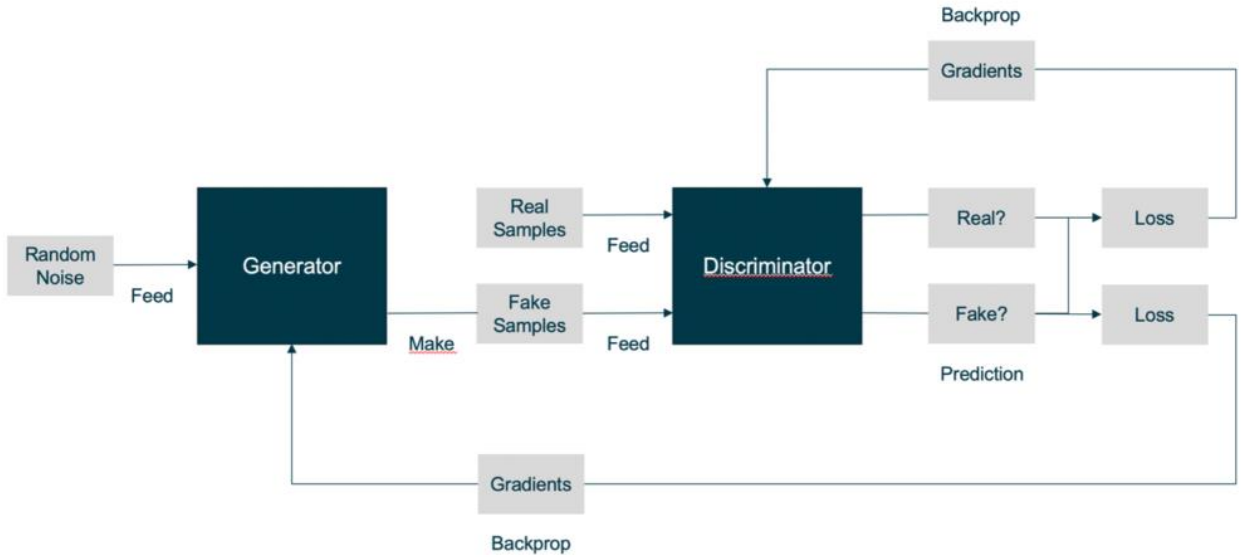
GAN-ların kibertəhlükəsizlik sahəsində istifadəsi ilə bağlı bir araşdırma şirkəti olan Cybersecurity Insiders tədqiqatına görə, müşahidə olunan 712 qərargahın 21%-i GAN-ları təhlükəsizlik tədqiqatları üçün kəşf etməkdədir.

Əlavə olaraq, GAN-ların kibertəhlükəsizlik üzrə istifadəsi ilə bağlı bir digər araşdırma şirkəti olan CSO Online'a əsasən, təhlükəsizlik mütəxəssislərinin 32%-i GAN-ları təhlükəsizlik tədqiqatlarında istifadə edəcəklərini planlaşdırırlar.

Bu statistikalar, GAN-ların kibertəhlükəsizlik sahəsində müvəffəqiyyətli şəkildə istifadə edilməsinin gələcəkdə daha da artacağını vurğulayır. GAN-ların tətbiqatı, məlumatların yaradılması, təhlili və təşkilatlar üçün potensial təhlükələrin aşkar edilməsində effektiv bir alətdir. Bu, təhlükəsizlik təşkilatlarının məhdudiyyətləri aşmaq, mövcud təhlükələrə daha sürətli reaksiya verərək, təhlükəsizlik standartlarını yüksəltmək və təhlükəsizlik strategiyalarını inkişaf etdirmək üçün möhkəm bir zəmin yaradır.

Tədqiqat nəticələrinə əsasən, istifadə olunan alqoritmlərin performansını qiymətləndirmək kritik bir mərhələdir. Bu bölüm, müşahidə edilən məlumatlar

əsasında, istifadə olunan alqoritmlərin performansının dəyərləndirilməsi, müqayisəli tədqiqatlar və nəticələrin ətraflı analizi ilə müşayiət olunur.



Şək. 4.2 Generative Adversarial Networks (GAN).

4.3. İstifadə olunan alqoritmlərin performansı və müqayisəsi

- İstifadə olunan alqoritmlərin performansının qiymətləndirilməsi

Klassifikasiya modellərinin performansını ölçmək üçün əsas metrikalar aşağıdakılardır:

Doğruluq (Accuracy): Toplam doğru təxminlərin faizi. Yüksək Doğruluq, alqoritmlərin effektivliyini göstərir. Doğruluq, bir sinifləndirmə modelinin nə qədər effektiv olduğunu ölçmək üçün ən əhəmiyyətli metriklərdən biridir. Bu metrik, modelin düzgün təxmin etdiyi nümunələrin faizini göstərir. Əgər bir modelin doğruluğu yüksəkdirsə, bu deməkdir ki, model məlumatları düzgün şəkildə sinifləndirməkdə effektivdir.

Doğruluq metriki, modelin düzgün təxmin etdiyi nümunələrin sayını, ümumi nümunələr sayına bölərək hesablanır. Misal üçün, 100 nümunədən ibarət bir datasetiniz var və modeliniz 80 nümunəni düzgün şəkildə sınıflandırırsa, Doğruluq faizi 80%

olacaq. Yüksək Doğruluq, modelin məlumatları səhv olmadan sınıflandırmaqda ne qədər effektiv olduğunu göstərir.

Doğruluq metrikinin effektivliyi, modelin məlumatları səhv olmadan sinifləndirmədə nə qədər uğurlu olduğunu və modelin performansını qiymətləndirmək üçün vacibdir. Ancak, yalnız Doğruluq metrikinə dayanmaq zaman zaman ədalətsiz nəticələrə səbəb ola bilər.

Həssaslıq (Precision). Pozitiv nümunələrin doğru təxmin edilə bilən faizi. Alqoritmlərin doğru məlumatları tapmaqda neçə dəqiq olduğunu ölçən əhəmiyyətli bir metrikadır. Trafik analogiyası ilə Həssaslıqi izah etmək mümkündür. Bir yol işarəsi kimi təsəvvür edin. Həssaslıq, bu işarəni doğru bir şəkildə tanıya bilən və ya tanıya bilməyən sürücülərin sayını göstərir. Bir sürücü bu işarəni düzgün tanıdıqda, bu "düzgün müsbət" (true positive) kimi hesablanır. Həmçinin, başqa bir sürücü bu işarəni yanlış anladıqda, bu isə "yalan müsbət" (false positive) olur. Həssaslıq, doğru tanıyan işarələrin (true positives) cəmi işarələrin (true positives + false positives) faizini göstərir.

Əgər bir yolda sadəcə bir neçə işarə olsun və sürücülərin əksəriyyəti doğru işarəni tanıyarsa, bu yolda yüksək Həssaslıq var deməkdir. Lakin, eyni zamanda, əgər sürücülərin bir hissəsi yanlış işarəni tanıyarsa, bu da sürücülər üçün ciddi bir problem yarada bilər. Həssaslığin yüksək olması, sürücülərin doğru qərar vermələrində daha çox güvənin olmasını təmin edir.

Rəqəmsal Həssaslıq (Recall). Bu məzmun pozitiv nümunələrin neçə faizinin düzgün bir şəkildə təxmin edilə biləcəyini göstərir. Əsasən, bu məzmun, məlumatların doğru bir şəkildə tanınmasını vurğular.

Misal üzərində düşünək. Bir sürücü yol işarələrinin yalnız bir hissəsini tanıyır və digər hissələrini gözləməyir. Rəqəmsal Həssaslıq, bu sürücünün doğru tanıdığı işarələrin faizini göstərir. Ehtimal ki, bir yolun yalnız bir hissəsində işarələr var və sürücü yalnız bu hissədəki işarələri tanıyır. Bu, sürücünün bu hissədəki bütün işarələri

tanıdığı və sürücünün yoldakı işarələrin tamamını tanıma ehtimalının olmadığını göstərən bir hal ola bilər.

F1 Skoru. F1 Skoru, Həssaslıq və Rəqəmsal Həssaslığın harmonik ortalamasıdır. Balanslaşdırılmış bir performans ölçüsüdür və modelin effektivliyini tamamlayır. F1 Skoru, məlumatların doğru təxmin edilməsinin və ya modelin effektivliyinin qiymətləndirilməsində çox əhəmiyyətli bir metrikdir. Bu metrik, özəlliklə, təhlükəsizlik təşkilatları və şəbəkə təhlükəsizliyi sahələrində, xüsusilə də şifrlənmiş trafikin klassifikasiyası məsələlərində istifadə olunur.

F1 Skoru, modelin həm də doğru nəticələri, həm də yanlış təxminləri dəyərləndirərək, modelin performansını balanslaşdırır. Yüksək bir F1 Skoru, modelin həm pozitiv nümunələri doğru təxmin etmədə, həm də məlumatları doğru bir şəkildə tanıdığını göstərir. Müqayisəli tədqiqatlar və nəticələrin analizi, şəbəkə təhlükəsizliyi sahəsində alqoritmlərin effektivliyini qiymətləndirmək üçün əsas rol oynayır. Bu bölüm, hər bir alqoritmin performansını ətraflı şəkildə təhlil edir və fərqlərin səbəblərini aydınlaşdırır. Müqayisəli tədqiqatlar əsasında, ən uyğun alqoritmlər müəyyənləşdirilir və ən yaxşı performans göstərən variantlar müəyyən edir [54].

4.4. Klassifikasiya alqoritmlərinin inkişafı üçün gələcəkdəki nailiyyətlər

Klassifikasiya alqoritmlərinin gələcəkdəki inkişafına dair potensial texnoloji inkişaflar, təlim alqoritmlərinin evolyutsiyası, şifrlənmiş trafikin klassifikasiyasında potensial inkişaflar və gələcəkdəki tədqiqat yönümləri və nailiyyətlərin proqnozlaşdırılması ilə bağlı ətraflı məlumatlar yer alır. Gələcəkdə texnologiya sahəsində potensial inkişaflar və trendlər ətraflı şəkildə nəzərdən keçirilir.

Kuantum Kompüterlər: Deep learning və kuantum kompüterlərinin birliyi, klassifikasiya alqoritmləri üçün yeni bir dövrün əsasını qoya bilər. Kuantum kompüterlər, qramlarınızın sahəsində ətraflı hesablama imkanları təmin edir, bu isə daha kompleks məlumatların dərin analizini və klassifikasiyasını mümkün edir. Misal

üçün, kuantum bitlər, daha çox sürət və effektivlik ilə daha geniş məlumat setlərini təhlil etmək və təhlükəli fəaliyyətləri müəyyən etmək imkanını verir.

Daha Güclü Alqoritmlər: Deep learning alqoritmlərinin daha da inkişafı, klassifikasiya tədqiqatlarında daha effektiv nəticələr əldə etmək üçün yeni imkanlar yaradır. Daha kompleks, adaptiv və öz-öyrənən alqoritmlər, daha geniş məlumat setlərində daha yaxşı performans göstərmək və təhlükəli fəaliyyətləri daha effektiv şəkildə tanımaq üçün istifadə oluna bilər. Məsələn, təhlükəsizlik agentlikləri üçün istifadə olunan yeni nəsil alqoritmlər, şəbəkədəki təhlükəli fəaliyyətləri daha effektiv şəkildə aşkar etməyə kömək edir.

Sürətli Hesablama Texnologiyaları: Klassifikasiya alqoritmlərinin sürətli hesablama texnologiyaları ilə birlikdə istifadəsi, təhlükəli trafik olaylarını daha sürətli və dəqiq şəkildə tanımağa imkan verir. Bulud hesablama platformalarının və digər sürətli hesablama sistemlərinin inkişafı, alqoritmlərin daha sürətli işləməsini və bu zaman çərçivəsində təhlükəli fəaliyyətləri tanımağını təmin edir.

Bu potensial texnoloji inkişafları, deep learning əsasında klassifikasiya alqoritmlərinin effektivliyini artırmaq və şəbəkə təhlükəsizliyi sahəsində yeni inkişafların qapısını açmaq üçün geniş imkanlar təmin edir. Bu, kibertəhlükəsizlikdə daha yaxşı müdafiə strategiyalarının inkişafını və daha effektiv təhlükəsizlik tədbirlərinin icrasını təmin edir.

Daha İrəli Teknologiyaların İstifadəsi: Gələcəkdəki tədqiqatlar, daha irəli və sürətli texnologiyaların şifrlənmiş trafikin klassifikasiyasında istifadəsini müzakirə edəcəkdir. Bu, kuantum kompüterlər, kvant neyron şəbəkələri, və s. kimi yeni və inkişaf etmiş texnologiyaların potensialinə diqqət yetirir.

Məlumat Kompleksləşdirilməsi: Gələcəkdəki tədqiqatlar, məlumatın diversifikasiyası və kompleksləşdirilməsi ilə bağlı olacaq. Bu, fərqli məlumat növlərinin (səs, şəkil, mətn və s.) bir araya gətirilməsi və məlumatların daha geniş spektrumunda klassifikasiya üsullarının tətbiq edilməsini təmin edəcəkdir.

Avtomatlaşdırılmış Proseslər: Gələcəkdəki tədqiqatlar, klassifikasiya proseslərinin avtomatikləşdirilməsi və öz-öyrənməli sistemlərin daha geniş yayılması ilə bağlı olacaq. Bu, məlumatların avtomatik şəkildə təhlil edilməsini və təhlükəsizlik təşkilatlarının zaman və ehtiyac olduğu anda effektiv reaksiya datalməsini təmin edəcəkdir.

Ətraflı Anormallıq Aşkar Edilməsi: Gələcəkdəki tədqiqatlar, klassifikasiya alqoritmlərinin daha çox anormallıqları aşkar etməsi və təhlükəsizlik təşkilatlarına daha çox səlahiyyət verən dəqiqlik və effektivlik səviyyələrində işləməsi ilə bağlı olacaqdır.

Mobil Klassifikasiya Alqoritmləri: Gələcəkdəki tədqiqatlar, mobil cihazlar üçün klassifikasiya alqoritmlərinin inkişafına diqqət yetirəcəkdir. Həmin cihazların, günümüzün hərəkətli işləmə dünyasında təhlükəsizlik tədbirlərinin əhəmiyyətini artırması nəticəsində, bu alqoritmlər mobil təhlükəsizlik tətbiqləri üçün əsaslı bir rol oynayacaq. Mobil cihazlar, gündəlik həyatımızın ayrılmaz bir hissəsi halına gəlməkdədir və bu, onları potensial hədələr halına gətirir. Bu səbəbdən, bu cihazların təhlükəsizliyinin təmin edilməsi, mobil klassifikasiya alqoritmlərinin inkişafı ilə bağlı əhəmiyyətli bir məsələdir.

Klassifikasiya alqoritmlərinin əsas rəhbərlik etdiyi bu istiqamətlər, şəbəkə təhlükəsizliyi sahəsində daha yüksək səviyyədə məlumat işləmə və müdafiə imkanlarının təmin edilməsini və ya istehsal edilməsini təmin edəcəkdir. Texnoloji inkişafların bu istiqamətlərdə tətbiqi, klassifikasiya alqoritmlərinin daha effektiv olmasına və bu vasitə ilə təhlükəsizlik tədbirlərinin daha sürətli, dəqiq və müvafiq reaksiya verərək qorunmasına imkan verəcəkdir. Bu, müəyyən trendlərin və texnoloji inkişafların başa çıxılması üçün strateji planlar hazırlayan tədqiqatçılar və endirimçilər üçün dəyərli bir resursdur [63].

NƏTİCƏ

1. Şifrələnmiş trafikin təsnifatı, şifrələnmiş məlumatların qruplaşdırılması və klassifikasiyasını həyata keçirərək məlumatların gizliliyini və təhlükəsizliyini təmin etmək üçün əhəmiyyətli bir təhlükəsizlik tədbiridir.

2. Şifrələnmiş trafik təsnifatında Deep Learning alqoritmləri, yüksək dəqiqlik və effektivlik təmin edir. Bu texnologiyalar kibertəhlükəsizlik sahəsində yeni imkanlar yaradır.

3. Dərin öyrənmə alqoritmlərindən (xüsusilə RNN, LSTM, CNN) təbii dil emalında mətnin ümumiləşdirilməsi, təsnifatı, sual-cavab kimi problemlərin həllində uğurla istifadə olunduğu müşahidə edilmişdir.

4. Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Autoencoders (AE) və Generative Adversarial Networks (GAN) kimi əhəmiyyətli Deep Learning alqoritmləri, şifrələnmiş trafikin klassifikasiyasında effektiv nəticələr verir.

5. Fərqli şəbəkə şərtləri, zamanla dəqiqliklər və düşmən hücumlarına qarşı möhkəmliyin artırılması kimi məsələlər üçün modellərin yenilənməsi və təkmilləşdirilməsi vacibdir.

6. Məlumatların etiketlənməsi və məlumat çatışmazlığı problemlərinin həll edilməsi üçün öz-özünə öyrənmə metodlarının tətbiqi tədqiq edilməlidir.

İSTİFADƏ EDİLMİŞ ƏDƏBİYYATLAR

1. Abdulkadir Ş., Banu D., Hasan H. B., Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme. Gazi Mühendislik Bilimleri Dergisi 2017, 3(3): 47-64 gmbd.gazipublishing.com

2. Agrawal A., Bhatia A., Bahuguna A., Tiwari K., Haribabu K., Vishwakarma D., Kaushik R., "A survey on analyzing encrypted network traffic of mobile devices," Int. J. Inf. Secur., vol. 21, no. 4, pp. 873–915, Aug. 2022.

3. Ali G., Ali Ç., Deep learning and machine learning based anomaly detection in internet of things environments. Journal of the Faculty of Engineering and Architecture of Gazi University 37:4 (2022) 1945-1956.

4. Atac C., Akleylek S., A Survey on Security Threats and Solutions in the Age of IoT, European Journal of Science and Technology, 15, 36-42, 2019.

5. Baccouche M., Mamalet F., Wolf C., Garcia C., Baskurt A., "Sequential Deep Learning for Human Action Recognition," Springer, Berlin, Heidelberg, pp. 29–39, 2011.

6. Bengio Y., "Learning Deep Architectures for AI," Found. trends® Mach. Learn., vol. 2, no. 1, pp. 1–127, 2009.

7. Bengio Y., Courville A., Vincent P., "Representation Learning: A Review and New Perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, 2013.

8. Booij T. M., Chiscop I., Meeuwissen E., Moustafa N., Den Hartog F. T. (2021). ToN_IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets. IEEE Internet of Things Journal, 9(1), 485-496.

9. Butt S.A., Arshad A., Martinez L.D., IoT Smart Health Security Threats Shariq, 2019 19th International Conference on Computational Science and Its Applications (ICCSA), 26-31, 2019.

10. Camelo M., Soto P., Latre S., ‘‘A general approach for traffic classification in wireless networks using deep learning,’’ IEEE Trans. Netw. Service Manage., vol. 19, no. 4, pp. 5044–5063, Dec. 2022.

11. Chicco D., Sadowski P., Baldi P., “Deep autoencoder neural networks for gene ontology annotation predictions,” in Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB ’14, 2014, pp. 533-540.

12. Cirean D. C., Meier U., Masci J., Gambardella L. M., “Flexible, High Performance Convolutional Neural Networks for Image Classification,” in Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, pp. 1237–1242, 2012.

13. Ciresan D. C., Meier U., Gambardella L. M., Schmidhuber J., “Convolutional Neural Network Committees for Handwritten Character Classification,” in 2011 International Conference on Document Analysis and Recognition, pp. 1135–1139, 2011.

14. Ciresan D., Giusti A., Gambardella L. M., Schmidhuber J., “Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images,” in Advances in neural information processing systems, pp. 2843–2851, 2012

15. Cireşan D., Meier U., Masci J., Schmidhuber J., “Multi-column deep neural network for traffic sign classification,” Neural Networks, vol. 32, pp. 333–338, 2012.

16. Cisco. (2019). Cisco Visual Networking Index: Forecast and Trends, 2017–2022. Retrieved from https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

17. Coates A., Ng A., Lee H., “An Analysis of Single-Layer Networks in Unsupervised Feature Learning,” in PMLR, 2011, pp. 215–223.

18. Collobert R., Weston J., “A unified architecture for natural language processing:Deep neural networks with multitask learning,” in Proceedings of the 25th

international conference on Machine learning - ICML ’08, vol. 20, no. 1, pp. 160–167, 2008.

19. Cook A., Mısırlı G., Fan Z., Anomaly Detection for IoT Time-Series Data, A Survey, IEEE Internet of Things Journal, 2019.

20. Deng L., Yu D., “Deep Learning: Methods and Applications,” Found. Trends® Signal Process., vol. 7, no. 3-4, pp. 197-387, 2014.

21. Dettmers Tim, “Deep Learning in a Nutshell: History and Training Parallel Forall,” 2015. [Online]. Available: https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-history-training/. [Accessed: 20- Mar-2017].

22. Erfani M., Shoeleh F., Dadkhah S., Kaur B., Xiong P., Iqbal S., Ghorbani, A. A. (2021, October). A feature exploration approach for IoT attack type classification. In 2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech) (pp. 582-588). IEEE.

23. Falcao A. X., Papa J. P. (Eds.). (2022). Optimum-Path Forest: Theory, Algorithms, and Applications. Academic Press. Ioannou, C., & Vassiliou, V. (2021). Network attack classification in IoT using support vector machines. Journal of sensor and actuator networks, 10(3), 58.

24. Farfade S. S., Saberian M., Li L. J., “Multiview Face Detection Using Deep Convolutional Neural Networks,” in Proceedings of the 5th ACM on International Conference on Multimedia Retrieval., pp. 643–650, 2015.

25. Gao J., Deng L., Gamon M., He X., “Modeling interestingness with deep neural networks,” 14/304, 863, 2014.

26. Gatys L. A., Ecker A. S., Bethge M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2414-2423).

27. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Bengio Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

28. Graves A., Mohamed A., Hinton G., "Speech recognition with deep recurrent neural networks," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649, 2013.

29. Hasan M., Islam. M.M., Zarif MI., Hashem M.M., Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches, Internet of Things 7, 2019.

30. Havaei M., "Brain tumor segmentation with Deep Neural Networks," Med. Image Anal., vol. 35, pp. 18-31, Jan. 2017.

31. He K., Zhang X., Ren S., Sun J., "Deep Residual Learning for Image Recognition," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

32. Hinton G. E., Salakhutdinov R. R., "Replicated Softmax: an Undirected Topic Model," in Advances in Neural Information Processing Systems 22 , 2009, pp. 1607–1614.

33. Islam U., Muhammad A., Mansoor R., Hossain M. S., Ahmad I., Eldin E. T., Shafiq M. (2022). Detection of distributed denial of service (DDoS) attacks in IOT based monitoring system of banking sector using machine learning models. Sustainability, 14(14), 8374.

34. Işın A., Direkoğlu C., Şah M., "Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods," Procedia Comput. Sci., vol. 102, pp. 317-324, 2016.

35. Jarjis A. H., Al Zubaidi N. Y. S., Pehlivanoglu, M. K. (2023). Cyber Attacks Classification on Enriching IoT Datasets. EAI Endorsed Transactions on Internet of Things, 9(3).

36. Kalıpcıoğlu K.C., Toğay C., Yolaçan E.N., Son Kullanıcılar İçin Anomali Saldırı Tespit Sistemleri, Journal of Engineering and Architecture Faculty of Eskisehir Osmangazi University, 27 (3), 199-212, 2019

37. Karmakar K.K., Varadharajan V., Tupakula U., Nepaly S., Thapa C., Towards a Security Enhanced Virtualised Network Infrastructure for Internet of Medical Things (IoMT), 6th IEEE International Conference on Network Softwarization, 257-261, 2020.

38. Karpathy A., Fei-Fei L., "Deep Visual1Semantic Alignments for Generating Image Descriptions," in CVPR, pp. 3128–3137, 2015.

39. Kilimci Z.H., Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (2), 635-650, 2019.

40. Kingma D. P., Ba J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

41. Kıyancı S., Mehmet Abi, A technological step in history education material: NFC, Research And Experience Journal (REJ), 4 (2), 2019.

42. Koroniotis N., Moustafa N., Sitnikova E., Turnbull B. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. Future Generation Computer Systems, 100, 779-796.

43. Kozik R., Pawlicki M., Choraś, M. (2021). A new method of hybrid time window embedding with transformer-based traffic data classification in IoT-networked environment. Pattern Analysis and Applications, 24(4), 1441-1449.

44. LeCun Y., Bengio Y., and Hinton G., "Deep learning," Nature, vol. 521, pp. 436-444, 2015.

45. Lee C. S., "Human vs. Computer Go: Review and Prospect [Discussion Forum]," IEEE Comput. Intell. Mag., vol. 11, no. 3, pp. 67–72, Aug. 2016.

46. Lounis K., Zulkernine M., Attacks and Defenses in Short-Range Wireless Technologies for IoT, IEEE Access, 8, 88892-88932, 2020.

47. Mısır O., Görkem L., Nesnelerin İnterneti için MQTT ile Hiyerarşik Haberleşme, Journal of New Results in Engineering and Natural Sciences, 2, 1-11, 2020.

48. Moustafa N., Slay J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). Military Communications and Information Systems Conference (MilCIS), 1-6.

49. Nascita A., Cerasuolo F., Di Monda D., Garcia J. T. A., Montieri A., Pescape, A. (2022, May). Machine and deep learning approaches for IoT attack classification. In IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 1-6). IEEE.

50. Nikravan M., Movaghar A., Hosseinzadeh M., A lightweight signcryption scheme for defense against fragment duplication attack in the 6LoWPAN networks, Peer-to-Peer Netw. Appl. Springer Science & Business Media, 12, 209–226, 2019.

51. Papadogiannaki E., Ioannidis S., ‘‘A survey on encrypted network traffic analysis applications, techniques, and countermeasures,’’ ACM Comput. Surv., vol. 54, no. 6, pp. 1–35, Jul. 2022.

52. Paul J., Yaacoub A., Noura M., Noura H.N., Ola Salman O., Yaacoub E., Couturier R., Chehab A., Securing internet of medical things systems: Limitations, issues and recommendations, Future Generation Computer Systems 105, 581–606, 2020

53. Prabu U., Geetha V., ‘‘Self-organizing deep learning model for network traffic classification,’’ in Proc. Inventive Commun. Comput. Technol. (ICICCT). Singapore: Springer, 2022, pp. 419–425.

54. Protogerou A., Papadopoulos S., Drosou A., Tzovaras D., Refanidis I., A graph neural network method for distributed anomaly detection in IoT, Evolving Systems 12, 19-36, 2021.

55. Rawat A. S., Wang Q. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), 2352-2449.

56. Ray S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.

57. Russakovsky O., “ImageNet Large Scale Visual Recognition Challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211-252, Dec. 2015.

58. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A. C., Fei-Fei L. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), 211-252.

59. Sahu A. K., Sharma S., Tanveer M., Raja, R. (2021). Internet of Things attack detection using hybrid Deep Learning Model. Computer Communications, 176, 146-154.

60. Sathyanarayana A.., “Sleep Quality Prediction From Wearable Data Using Deep Learning,” JMIR mHealth uHealth, vol. 4, no. 4, p. e125, Nov. 2016.

61. Sboev A., Litvinova T., Gudovskikh D., Rybka R., Moloshnikov I., “Machine Learning Models of Text Categorization by Author Gender Using Topic-independent Features,” Procedia Comput. Sci., vol. 101, pp. 135–142, 2016.

62. Schmidhuber J., “Deep learning in neural networks: An overview,” Neural Networks, vol. 61, pp. 85–117, 2015.

63. Schuster M., Paliwal K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673-2681.

64. Shahraki A., Abbasi M., Taherkordi A., Jurcut A. D., ‘‘Active learning for network traffic classification: A technical study,’’ IEEE Trans. Cogn. Commun. Netw., vol. 8, no. 1, pp. 422–439, Mar. 2022.

65. Shen Y., He X., Gao J., Deng L., Mesnil G., “Learning semantic representations using convolutional neural networks for web search,” in Proceedings of the 23rd International Conference on World Wide Web - WWW ’14 Companion, 2014, pp. 373–374.

66. Song H. A., Lee S. Y., “Hierarchical Representation Using NMF,” in International Conference on Neural Information Processing., pp. 466–473, 2013.

67. Swarna Priya R.M., Maddikunta P.K.R., Parimala M., Koppu S., Gadekallu T.R., Chowdhary C.L., Alazab M., An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture, Computer Communications 160, 139-149, 2020.

68. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Rabinovich A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

69. Tan M., Le Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, PMLR 97:6105-6114.

70. Toraman S., Türkoğlu İ., A new method for classifying colon cancer patients and healthy people from FTIR signals using wavelet transform and machine learning techniques, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (2), 933-942, 2019.

71. Ullah I., Mahmoud Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. IEEE Access, 9, 103906-103926.

72. Vigoya L., Fernandez D., Carneiro V., Cacheda F., Annotated Dataset for Anomaly Detection in a Data Center with IoT Sensors, 20 (13), 3745, 2020.

73. Wang S., Cao J., Yu P. S., ‘‘Deep learning for spatio-temporal data mining: A survey,’’ IEEE Trans. Knowl. Data Eng., vol. 34, no. 8, pp. 3681–3700, Aug. 2022.

74. Wozniak, M., Silka J., Wieczorek M., Alrashoud M. (2020). Recurrent neural network model for IoT and networking malware threat detection. IEEE Transactions on Industrial Informatics, 17(8), 5583-5594.

75. Yadav S., Subramanian S., “Detection of Application Layer DDoS attack by feature learning using Stacked AutoEncoder,” in International Conference on

Computational Techniques in Information and Communication Technologies, pp. 361–366, 2016.

76. You L., Li Y., Wang Y., Zhang J., Yang Y., “A deep learning-based RNNs model for automatic security audit of short messages,” in 2016 16th International Symposium on Communications and Information Technologies (ISCIT), pp. 225– 229, 2016.

77. Yousefi-Azar M., Hamey L., “Text Summarization Using Unsupervised Deep Learning,” Expert Syst. Appl., vol. 68, pp. 93–105, Feb. 2017.

78. Zhang X. Y., Wu Q., Zhu Y., Zhang R. (2020). Deep learning for intrusion detection: opportunities and challenges. IEEE Network, 34(6), 122-128.

79. Zhang X. Y., Zhu Y., Leung H. (2019). Deep Learning in Network Security. In Deep Learning in Big Data Analytics (pp. 147-176). Springer, Cham.

80. Zhang Q., “Deep learning based classification of breast tumors with shear-wave elastography,” Ultrasonics, vol. 72, pp. 150-157, 2016.

81. Zhou S., Chen Q., Wang X., “Active deep learning method for semi-supervised sentiment classification,” Neurocomputing, vol. 120, pp. 536– 546, 2013.

82. Zhou S., Chen Q., Wang X., “Fuzzy deep belief networks for semi-supervised sentiment classification,” Neurocomputing, vol. 131, pp. 312– 322, 2014.